

BAYESIAN METHODS IN THE SHAPE INVARIANT MODEL (I): POSTERIOR CONTRACTION RATES ON PROBABILITY MEASURES

BY DOMINIQUE BONTEMPS* AND SÉBASTIEN GADAT*

Institut Mathématiques de Toulouse, Université Paul Sabatier

In this paper, we consider the so-called Shape Invariant Model which stands for the estimation of a function f^0 submitted to a random translation of law g^0 in a white noise model. We are interested in such a model when the law of the deformations is *unknown*. We aim to recover the law of the process \mathbb{P}_{f^0, g^0} .

In this perspective, we adopt a Bayesian point of view and find prior on f and g such that the posterior distribution concentrates at a polynomial rate around \mathbb{P}_{f^0, g^0} when n goes to $+\infty$. We intensively use some Bayesian non parametric tools coupled with mixture models and believe that some of our results obtained on this mixture framework may be also of interest for frequentist point of view.

1. Introduction. We are interested in this work in the so-called Shape Invariant Model (SIM). Such model aims to describe a statistical process which involves a deformation of a functional shape according to some randomized geometric variability. Such geometric deformation of a common unknown shape may be well-suited in various and numerous fields, like image processing (see for instance [AGP91] or [PMRC10]). It corresponds to a particular case of the general Grenander's theory of shapes (see [GM07] for a detailed introduction on this topic). This kind of model is also useful in medicine: the recent work of [Big11] deals with the differentiation between normal and arrhythmic cycles in electrocardiogram. It appears in genetics if one deals with some delayed activation curves of genes when drugs are administrated to patients, or in Chip-Seq estimation when translations in protein fixation yield randomly shifted counting processes (see for instance [MMW07] and [BGKM12]). It also occurs in econometric for the analysis of Engel curves [BCK07], in landmark registration [Big06]. . .

Such a model has received a large interest in the statistical community as pointed by the large amount of references on this subject. Some works

*The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under references ANR-JCJC-SIMI1 DEMOS and ANR Bandhits.

AMS 2000 subject classifications: Primary 62G05, 62F15; secondary 62G20

Keywords and phrases: Grenander's pattern theory, Shape Invariant Model, Bayesian methods, Convergence rate of posterior distribution, Non parametric estimation

consider a semi-parametric approach for the estimation (*self-modeling regression framework* used by [KG88] and [BGV09]). In [Cas12], the author applies some Bayesian techniques to obtain also statistical results on SIM in a semi-parametric setting when the level of noise on observations asymptotically vanishes. Older approaches use parametric settings (see [GM01] and the discussion therein for an overview) and study the so-called Fréchet mean of pattern. Standard M -estimation or Bayesian methods are exploited in [BGL09] or [AAT07] and same authors develop in [AKT10] a nice stochastic algorithm to run estimation in such a model. Some recent works follow some testing strategies to obtain curve registration [CD11], [Col12]. At last, note that [BG10] obtains some minimax adaptive results for non-parametric estimations in the Shape Invariant Model when one knows the law of the randomized translations.

All these works are interested in the statistical process of deformation of the "mean common shape" and generally aim to recover this unknown functional object according to noisy i.i.d. observations. Moreover, the Shape Invariant Model is considered as a standard benchmark for statistical methods which aim to compute estimations in some more general deformable models. Of course, the SIM could be extended to some more general situations of geometrical deformations described through an action of a finite dimensional Lie Group (see [BCG12] for a precise non parametric description). We have decided to restrict our work here to the simplest case of the one dimensional Lie group of translation \mathbb{S}^1 to warp the functional objects.

This work has been inspired by several discussions with Alain Trouvé about the work [AKT10] for the study of the Shape Invariant Model. We aim to extend their parametric Bayesian framework to the non-parametric setting and then study the behaviour of some posterior distributions. Hence, the motivation of the paper is mainly theoretical: we want to describe the asymptotic evolution of the posterior probability distributions when data are coming from the SIM. Of course, we need to build suitable prior which yield nice contraction rate for this posterior distribution. We have decided to consider the general case where both the functional shape and the probability distribution of the deformations are unknown. Indeed, it corresponds to the more realistic case. From the best of our knowledge, no sharp statistical results have been derived yet in this non-parametric situation.

Our work will describe the evolution of the posterior distribution when the number of observations grows to $+\infty$ with a fixed noise level σ . It is an important difference with the study of the asymptotically vanishing noise situation ($\sigma \rightarrow 0$). It is itself a special feature of the Shape Invariant Model: there is no obvious Le Cam equivalence of experiments (see [LCY00]) for the

SIM between the experiments when $n \mapsto +\infty$ and when $\sigma \mapsto 0$. It is illustrated by the very different minimax results obtained in [BG10] ($n \mapsto +\infty$) and in [BG12] ($\sigma \mapsto 0$). We will use in the sequel quite standard Bayesian non parametric methods to obtain the frequentist consistency and some contraction rates of the Bayesian procedures. Such tools rely on some important contributions of [BSW99] and [GGvdV00] for the posterior behaviour in general situations, as well as Bayesian properties on mixture models stated in [GvdV01] and [GW00].

The paper is organised as follows. Section 2 presents a sharp description of the Shape Invariant Model (shortened as SIM in the sequel), as well as standard elements on Bayesian and Fourier analysis. It also provides some notations for mixture models. It ends with the statement of the posterior contraction around the true law on functional curves, which is our main result. Section 3 provides a metric description of the important probability spaces of the model. At last, Section 4 presents the proof of this main result. We end the paper with numerous challenging issues.

We gather in the appendix sections some technical points: the metric description of the Shape Invariant Model embedded in a special randomized curves space and the calibration of suitable priors for the SIM.

2. Model, notations and main results.

2.1. Statistical settings.

Shape Invariant Model. We recall here the random Shape Invariant Model. We assume f^0 to be a function which belongs to a subset \mathcal{F} of smooth functions. We also consider a probability measure g^0 which is an element of the set $\mathfrak{M}([0, 1])$. This last set stands for the set of probability measures on $[0, 1]$. We observe n realizations of noisy and randomly shifted complex valued curves Y_1, \dots, Y_n coming from the following white noise model

$$(2.1) \quad \forall x \in [0, 1] \quad \forall j = 1 \dots n \quad dY_j(x) := f^0(x - \tau_j)dx + \sigma dW_j(x).$$

Here, f^0 is the *mean* pattern of the curves Y_1, \dots, Y_n although the random shifts $(\tau_j)_{j=1 \dots n}$ are sampled independently according to the probability measure g^0 . Moreover, $(W_j)_{j=1 \dots n}$ are independent complex standard Brownian motions on $[0, 1]$ and model the presence of noise in the observations, the noise level is kept fixed in our study and is set to 1 for sake of simplicity.

In the sequel, $f^{-\tau}$ will denote the pattern f shifted by τ , that is to say the function $x \mapsto f(x - \tau)$. Complex valued curves are considered here for the simplicity of notations. However all our results can be adapted to the

simpler case where all curves Y_j 's are real valued. A complex standard Brownian motion W_t on $[0, 1]$ is such that W_1 is a standard complex Gaussian random variable, whose distribution is denoted by $\mathcal{N}_{\mathbb{C}}(0, 1)$; a standard complex Gaussian random variable have independent real and imaginary parts with a real centered Gaussian distribution of variance $1/2$.

This work will address the question of the behaviour of some posterior distributions on $\mathcal{F} \otimes \mathfrak{M}([0, 1])$ given some functional n -sample (Y_1, \dots, Y_n) . Since our work will be mainly asymptotic with $n \rightarrow +\infty$, we intensively use some standard notation such as " \lesssim " which refers to an inequality up to a multiplicative absolute constant. In the meantime, $a \sim b$ stands for $a/b \rightarrow 1$.

Bayesian framework. Since most of statistical works on the SIM are frequentists, we have decided to briefly recall here the Bayesian formalism following the presentation of [GGvdV00]. Familiar readers can thus omit this paragraph.

Functional objects f^0 and g^0 we are looking for, belong to $\mathcal{F} \otimes \mathfrak{M}([0, 1])$ and for any couple $(f, g) \in \mathcal{F} \otimes \mathfrak{M}([0, 1])$, equation (2.1) describes the law of one continuous curve. Its law is denoted $\mathbb{P}_{f,g}$ and possesses a density $p_{f,g}$ with respect to the Wiener measure on the sample space. Since f^0 and g^0 are unknown, \mathbb{P}_{f^0,g^0} is also unavailable but belongs to a set \mathcal{P} of probability measure over the sample space. This set \mathcal{P} is the set of all possible measures described by (2.1) when (f, g) varies into $\mathcal{F} \otimes \mathfrak{M}([0, 1])$.

Given some prior distribution Π_n on \mathcal{P} (generally defined through a prior on $\mathcal{F} \otimes \mathfrak{M}([0, 1])$), Bayesian procedures are generally built using the posterior distribution defined by

$$\Pi_n(B|Y_1, \dots, Y_n) = \frac{\int_B \prod_{j=1}^n p(Y_j) d\Pi_n(p)}{\int_{\mathcal{P}} \prod_{j=1}^n p(Y_j) d\Pi_n(p)},$$

which is a random measure on \mathcal{P} that depends on the observations Y_1, \dots, Y_n . For instance, Bayesian estimators can be obtained using the mode, the mean or the median of the posterior distribution. This is exactly the approach adopted by [AKT10] which is mainly dedicated to compute such a posterior mean in a parametric setting with a stochastic EM algorithm.

The posterior distribution is then said *consistent* if it concentrates to arbitrarily small neighbourhoods of \mathbb{P}_{f^0,g^0} in \mathcal{P} with a probability tending to 1 when n grows to $+\infty$. One *frequentist* property of such a posterior distribution describes the contraction rate of such neighbourhoods meanwhile still capturing most of posterior mass. According to equation (2.1), we thus tackle such a Bayesian consistency and compute such convergence rates in

the frequentist paradigm. Of course, these properties will highly depend on the metric structure of the sets \mathcal{P} and \mathcal{F} .

Functional setting and Fourier analysis. Without loss of generality, the function f^0 is assumed to be periodic with period 1 and to belong to a subset \mathcal{F} of $L^2_{\mathbb{C}}([0, 1])$, the space of squared integrable functions on $[0, 1]$ endowed with the euclidean norm $\|h\| := \int_0^1 |h(s)|^2 ds$. Moreover, each element $h \in L^2_{\mathbb{C}}([0, 1])$ may naturally be extended to a periodic function on \mathbb{R} of period 1. Since we will intensively use some Fourier analysis in the sequel, let us first recall some notations: i will stand for the complex number such that $i^2 = -1$. The Fourier coefficients of h are denoted

$$(2.2) \quad \theta_{\ell}(h) := \int_0^1 e^{-i2\pi\ell t} h(t) dt.$$

All along the paper, we will often use the parametrisation of any element of $h \in L^2_{\mathbb{C}}([0, 1])$ through its Fourier expansion and will simply use the notation $(\theta_{\ell})_{\ell \in \mathbb{Z}}$ instead of $(\theta_{\ell}(h))_{\ell \in \mathbb{Z}}$.

Our work is dedicated to the analysis of SIM when \mathcal{F} models smooth functions of $[0, 1]$. Hence, natural subspaces of $L^2_{\mathbb{C}}([0, 1])$ are Sobolev spaces \mathcal{H}_s with a smoothness parameter s :

$$\mathcal{H}_s := \left\{ f \in L^2_{\mathbb{C}}([0, 1]) \mid \sum_{\ell \in \mathbb{Z}} (1 + |\ell|^{2s}) |\theta_{\ell}(f)|^2 < +\infty \right\}.$$

In the sequel, we aim to find prior on \mathcal{P} that reaches good frequentist properties, and if possible adaptive with the smoothness parameter s since this parameter is generally unknown. We will consider only some regular cases when $s \geq 1$, the quantity $\sum_{\ell} \ell^2 |\theta_{\ell}|^2$ is thus bounded and we denote the Sobolev norm

$$\|\theta\|_{\mathcal{H}_1} := \sqrt{\sum_{\ell \in \mathbb{Z}} \ell^2 |\theta_{\ell}|^2}.$$

It will also be useful to consider in some cases Fourier "thresholded" elements of \mathcal{H}_s . Hence, we set for any integer ℓ (which is the frequency threshold)

$$\mathcal{H}^{\ell} := \{ f \in L^2_{\mathbb{C}}([0, 1]) \mid \forall |k| > \ell \quad \theta_k(f) = 0 \}.$$

Mixture model. According to equation (2.1), we can write in the Fourier domain that

$$\forall \ell \in \mathbb{Z} \quad \forall j \in \{1 \dots n\} \quad \theta_{\ell}(Y_j) = \theta_{\ell}^0 e^{-i2\pi j \tau_j} + \xi_{\ell, j},$$

where $\theta^0 := (\theta_\ell^0)_{\ell \in \mathbb{Z}}$ denotes the true unknown Fourier coefficients of f^0 . Owing to the white noise model, the variables $(\xi_{\ell,j})_{\ell,j}$ are independent standard (complex) Gaussian random variables: $\xi_{\ell,j} \sim_{i.i.d.} \mathcal{N}_{\mathbb{C}}(0, 1), \forall \ell, j$.

For sake of simplicity, γ will refer to $\gamma(z) := \pi^{-1}e^{-|z|^2}, \forall z \in \mathbb{C}$, the density of the standard complex Gaussian centered distribution $\mathcal{N}_{\mathbb{C}}(0, 1)$, and $\gamma_\mu(\cdot) := \gamma(\cdot - \mu)$ is the density of the standard complex Gaussian with mean μ . We keep also the same notation for p dimensional complex Gaussian densities $\gamma(z) := \pi^{-p}e^{-\|z\|^2}, \forall z \in \mathbb{C}^p$, where $\|z\|$ is the euclidean p dimensional norm of the complex vector z .

For any frequency ℓ , equation (2.1) implies that $\theta_\ell(Y)$ follows a mixture of complex Gaussian standard variables with mean $\theta_\ell^0 e^{-i2\pi\ell\varphi}, \varphi \in [0, 1]$:

$$\theta_\ell(Y) \sim \int_0^1 \gamma_{\theta_\ell^0 e^{-i2\pi\ell\varphi}}(\cdot) dg(\varphi).$$

In the sequel, for any phase $\varphi \in [0, 1]$ sampled according to any distribution g , and for any $\theta \in \ell^2(\mathbb{Z})$, $\theta \bullet \varphi$ will denote the element of $\ell^2(\mathbb{Z})$ given by

$$\forall \ell \in \mathbb{Z} \quad (\theta \bullet \varphi)_\ell := \theta_\ell e^{-i2\pi\ell\varphi}.$$

When θ is a complex vector, for instance $\theta = (\theta_{-\ell}, \dots, \theta_\ell)$, we keep the same notation $\theta \bullet \varphi := (\theta_{-\ell} e^{i2\pi\ell\varphi}, \dots, \theta_0, \theta_1 e^{-i2\pi\varphi}, \dots, \theta_\ell e^{-i2\pi\ell\varphi})$ to refer to the $2\ell + 1$ dimensional vector. It corresponds to a rotation of each coefficient θ_ℓ around the origin with an angle $2\pi\ell\varphi$. According to this notation, the law of the infinite series (of Fourier coefficients of Y) can thus be rewritten as

$$\theta(Y) \sim \int_0^1 \gamma_{\theta \bullet \varphi}(\cdot) dg(\varphi).$$

One should remark the important fact that from one frequency to another, the rotations used to build $\theta(Y)$ are not independent, which traduces the fact that the coefficients $(\theta_\ell(Y))_\ell$ are highly correlated.

2.2. Notations on Mixture models. Our study will intensively use some classical tools of mixture models, see for instance the papers of [GvdV01] or [GW00]. We thus choose to keep some notations already used in such works.

For any vector $\theta \in \ell_{\mathbb{C}}^2(\mathbb{Z})$ corresponds a function $f \in L^2([0, 1])$ according to equation (2.2) and for any measure $g \in \mathfrak{M}([0, 1])$, $\mathbb{P}_{\theta,g}$ will refer to the law of the vector of $\ell^2(\mathbb{Z})$ described by the location mixture of Gaussian variables:

$$\mathbb{P}_{\theta,g} := \int_0^1 \gamma_{\theta \bullet \varphi}(\cdot) dg(\varphi).$$

This mixture model is of infinite dimension since θ belongs to $\ell^2(\mathbb{Z})$. Following an obvious notation shortcut, $\mathbb{P}_{f,g}$ will be its equivalent for the functional law on curves derived from $\mathbb{P}_{\theta,g}$. When θ is of finite length k , $p_{\theta,g}$ will be the density with respect to the Lebesgue measure on \mathbb{C}^k of the law $\mathbb{P}_{\theta,g}$:

$$\forall z \in \mathbb{C}^k \quad p_{\theta,g}(z) := \int_0^1 \gamma(z - \theta \bullet \varphi) dg(\varphi).$$

We also use standard objects such as the Hellinger distance d_H between probability measures and the Total Variation distance d_{TV} , as well as covering numbers of metric spaces such as $D(\epsilon, \mathcal{P}, d)$. These objects are precisely described in Appendix A.

Bayesian frequentist consistency rate. In our setting, d is chosen according to one of the metric introduced above (d_H or d_{TV}) on the set

$$\mathcal{P} := \{\mathbb{P}_{f,g} | (f, g) \in \mathcal{H}_s \otimes \mathfrak{M}([0, 1])\}.$$

We can now remind Theorem 2.1 of [GGvdV00] which will be useful for our purpose.

THEOREM 2.1 (Posterior consistency and convergence rate, [GGvdV00]). *Assume that a sequence $(\epsilon_n)_n$ with $\epsilon_n \rightarrow 0$ and $n\epsilon_n^2 \rightarrow +\infty$, a constant $C > 0$, and a sequence of sets $\mathcal{P}_n \subset \mathcal{P}$ satisfy*

$$(2.3) \quad \log D(\epsilon_n, \mathcal{P}_n, d) \leq n\epsilon_n^2$$

$$(2.4) \quad \Pi_n(\mathcal{P} \setminus \mathcal{P}_n) \leq e^{-n\epsilon_n^2(C+4)}$$

$$(2.5) \quad \Pi_n(\mathbb{P}_{f,g} \in \mathcal{P} | d_{KL}(\mathbb{P}_{f^0,g^0}, \mathbb{P}_{f,g}) \leq \epsilon_n^2, V(\mathbb{P}_{f^0,g^0}, \mathbb{P}_{f,g}) \leq \epsilon_n^2) \geq e^{-n\epsilon_n^2 C}.$$

Then there exists a sufficiently large M such that

$$\Pi_n(\mathbb{P}_{f,g} : d(\mathbb{P}_{f^0,g^0}, \mathbb{P}_{f,g}) \geq M\epsilon_n | Y_1, \dots, Y_n) \longrightarrow 0$$

in \mathbb{P}_{f^0,g^0} probability as $n \longrightarrow +\infty$.

The posterior concentration rate obtained in the above result is ϵ_n . The growing set \mathcal{P}_n is referred to as a Sieve over \mathcal{P} . Generally, this rate ϵ_n can be compared to the classical frequentist benchmark: for instance [GGvdV00] obtained for the Log Spline model a contraction rate $\epsilon_n = n^{-s/(2s+1)}$ when

the unknown underlying density belongs to an Hölder class $\mathcal{C}^s([0, 1])$, and this rate is known to be the optimal one (in the sense that it is the minimax one) in the frequentist paradigm over Hölder densities of regularity s (see [IH81]). Similarly, the recent work of [RR12] considers the situation of density estimation for infinite dimensional exponential families and reaches also contraction rates close or equal to the known optimal frequentist one.

2.3. Bayesian prior and posterior concentration in the randomly shifted curves model. We detail here the Bayesian prior Π_n on \mathcal{P} used to obtain a polynomial concentration rate. Note that such prior will be in our work independent on the unknown smoothness parameter s . As pointed in the paragraph above, it is sufficient to define some prior on the space $\mathcal{H}_s \otimes \mathfrak{M}([0, 1])$ since equation (2.1) will then transport this prior to a law Π_n on \mathcal{P} . The two parameters f and g are picked independently at random following the next prior distributions.

Prior on f . The prior on f is slightly adapted from [RR12]. It is defined on \mathcal{H}_s through

$$\pi := \sum_{\ell \geq 1} \lambda(\ell) \pi_\ell.$$

Given any integer ℓ , the idea is to decide to randomly switch on with probability $\lambda(\ell)$ all the Fourier frequencies from $-\ell$ to $+\ell$. Then, π_ℓ is a distribution defined on $\ell^2(\mathbb{Z})$ such that $\pi_\ell := \otimes_{k \in \mathbb{Z}} \pi_\ell^k$ and

$$\forall k \in \mathbb{Z} \quad \pi_\ell^k = 1_{|k| > \ell} \delta_0 + 1_{|k| \leq \ell} \mathcal{N}_{\mathbb{C}}(0, \xi_n^2).$$

The randomisation of selected frequencies is done using λ , a probability distribution on \mathbb{N}^* which satisfies for $\rho \in (1, 2)$:

$$\exists (c_1, c_2) \in \mathbb{R}_+ \quad \forall \ell \in \mathbb{N}^* \quad e^{-c_1 \ell^2 \log^\rho \ell} \lesssim \lambda(\ell) \lesssim e^{-c_2 \ell^2 \log^\rho \ell}.$$

The prior π depends on the variance of the Gaussian laws ξ_n used to sample the Fourier coefficients. In the sequel, we use a variance that depends on n according to

$$(2.6) \quad \xi_n^2 := n^{-\mu_s} (\log n)^{-\zeta},$$

where μ_s and ζ are parameters that may depend on s (non adaptive prior) or not (adaptive prior).

Prior on g . As our model does not seem so far from a mixture Gaussian model, a natural prior on g is built according to a Dirichlet process following the ideas of [GvdV01]. Given any finite base measure α that has a positive continuous density on $[0, 1]$ w.r.t. the Lebesgue measure, the Dirichlet process D_α generates a random probability measure g on $[0, 1]$. For any finite partition (A_1, \dots, A_k) of $[0, 1]$, the probability vector $(g(A_1), \dots, g(A_k))$ on the k -dimensional simplex has a Dirichlet distribution $Dir(\alpha(A_1), \dots, \alpha(A_k))$. Such process may be built according to the Stick-Breaking construction (see for instance [Fer73]).

2.4. Main result. Using the prior defined above, we obtain the following theorem on the randomly SIM.

THEOREM 2.2. *Assume that $f^0 \in \mathcal{H}_s$ with $s \geq 1$, then the values $\mu_s = 2/(2s + 2)$ and $\zeta = 0$ in the definition of ξ_n yield a non adaptive prior such that*

$$\Pi_n \{ \mathbb{P}_{f,g} \text{ s.t. } d_H(\mathbb{P}_{f,g}, \mathbb{P}_{f^0,g^0}) \leq M\epsilon_n | Y_1, \dots, Y_n \} = 1 + \mathcal{O}_{\mathbb{P}_{f^0,g^0}}(1)$$

when $n \rightarrow +\infty$, for a sufficiently large constant M such that. Moreover, the contraction rate ϵ_n is given by

$$\epsilon_n = n^{-s/(2s+2)} \log n.$$

The values $\mu = 1/4$ and $\zeta = 3/2$ yield the contraction rate

$$\Pi_n \{ \mathbb{P}_{f,g} \text{ s.t. } d_H(\mathbb{P}_{f,g}, \mathbb{P}_{f^0,g^0}) \leq M\epsilon_n | Y_1, \dots, Y_n \} = 1 + \mathcal{O}_{\mathbb{P}_{f^0,g^0}}(1)$$

for a sufficiently large constant M , when $n \rightarrow +\infty$ with

$$\epsilon_n = \begin{cases} n^{-s/(2s+2)} \log n & \text{if } s \in [1, 3] \\ n^{-3/8} \log n & \text{if } s \geq 3. \end{cases}$$

Let us briefly comment this result. It first describes the posterior concentration around some neighbourhood of the true law \mathbb{P}_{f^0,g^0} within a polynomial rate. Our prior is adaptive with the regularity s as soon as $s \in [1, 3]$ setting $\xi_n^2 = n^{-1/4}(\log n)^{-3/2}$. For this range of s , the convergence rate is $n^{-s/(2s+2)}$ up to a logarithmic term. To the best of our knowledge, the minimax frequentist rate is unknown for the problem on recovering \mathbb{P}_{f^0,g^0} when both f^0 and g^0 are unknown. An interpretation of such polynomial rate is rather difficult to provide. It may be interpreted as $-s/(2s + d)$ where d is the number of dimension to estimate in the model (f^0 and g^0). When s becomes larger than 3, the rate of Theorem (2.2) is "blocked" to $3/8$ (which

corresponds to $s/(2s+2)$ when $s=3$) and does not match with $s/(2s+2)$. This difficulty is mainly due to the important condition $w_\epsilon^2 \lesssim l_\epsilon$ in Theorem 3.1.

At last, the non adaptive prior based on $\xi_n^2 = n^{-2/(2s+2)}$ recovers the good rate $-s/(2s+2)$ for all s larger than 1.

The former result establish a result on the law $\mathbb{P}_{f,g} \in \mathcal{P}$. It is also possible to derive a second result on the objects $f \in \mathcal{H}_s$ themselves. This result is studied in [BG13] and provides a somewhat quite weak result on the posterior convergence towards the true objects f^0 and g^0 .

3. Metric description of the model. We aim to check conditions (2.4) and (2.5) and then apply Theorem 2.1. In this view, we first define in section 3.1 a sieve $\mathcal{P}_{\ell_\epsilon, w_\epsilon}$, and our goal is to find some optimal calibration of ϵ , l_ϵ and w_ϵ with respect to n . We thus need to find a lower bound of the prior mass around some Kullback-Leibler neighbourhood of $\mathbb{P}_{f^0, g^0} \in \mathcal{P}$. These sets are defined as

$$\mathcal{V}_{\epsilon_n}(\mathbb{P}_{f^0, g^0}, d_{KL}) = \{\mathbb{P}_{f,g} \in \mathcal{P} | d_{KL}(\mathbb{P}_{f^0, g^0}, \mathbb{P}_{f,g}) \leq \epsilon_n^2, V(\mathbb{P}_{f^0, g^0}, \mathbb{P}_{f,g}) \leq \epsilon_n^2\}.$$

This will be done indeed considering Hellinger neighbourhoods instead of Kullback-Leibler ones. A link between these two kinds of neighbourhood is given in section 3.2. In section 3.3, we work with the Hellinger neighbourhoods to exhibit some admissible sizes for ϵ_n , ℓ_n and w_n . At last, we prove Theorem 2.2 in section 4.1.

In all this section, we delay most technical proofs to the Appendix.

3.1. Entropy estimates. We first establish some useful results on the complexity of our model $\mathbb{P}_{f,g}$ when $f \in \mathcal{H}_s$ and $g \in \mathfrak{M}([0, 1])$ in various situations (f known, unknown, parametric or not).

3.1.1. Case of known f . We first give some useful results when f is known and belongs to a finite dimensional vector space (the number of active Fourier coefficients is restricted to $[-\ell, \ell]$ for a given ℓ). Then ℓ will be allowed to grow with n and depend on a parameter ϵ introduced below. Hence, f is described by the parameter $\theta = (\theta_{-\ell}, \dots, \theta_0, \dots, \theta_\ell)$, and we define the set of all possible Gaussian measures

$$\mathcal{A}_\theta := \{\gamma_{\theta \bullet \varphi}, \varphi \in [0, 1]\}.$$

Following the arguments of [GW00], it is possible to establish the following preliminary result.

PROPOSITION 3.1. *For any sequence $\theta \in \mathbb{C}^{2\ell+1}$, one has*

$$N_{\square}(\epsilon, \mathcal{A}_{\theta}, d_H) \leq \frac{4\pi\sqrt{2(2\ell+1)}\|\theta\|_{\mathcal{H}_1}}{\epsilon}(1 + o(1)),$$

where $o(1)$ goes to zero independently on ℓ and θ as $\epsilon \rightarrow 0$, and

$$\log N(\epsilon, \mathcal{A}_{\theta}, d_H) \lesssim \log \ell + \log \|\theta\|_{\mathcal{H}_1} + \log \frac{1}{\epsilon}.$$

Assume now that g possesses a finite number of k points in its support, one can deduce from the proposition above a simple corollary that exploits the complexity of the simplex of dimension $k - 1$ (see for instance the proof of Lemma 2 in [GW00]).

PROPOSITION 3.2. *Assume that f is parametric and known ($\theta \in \mathbb{C}^{2\ell+1}$) and define*

$$\mathcal{M}_{\theta}^k := \left\{ \sum_{i=1}^k g(\varphi_i) \gamma_{\theta \bullet \varphi_i} : \varphi_i \in [0, 1], g(\varphi_i) \geq 0, \forall i \in \llbracket 1, k \rrbracket \text{ and } \sum_{i=1}^k g(\varphi_i) = 1 \right\}$$

for a number of components k that may depend on ϵ (as ℓ does). Then

$$H_{\square}(\epsilon, \mathcal{M}_{\theta}^k, d_H) \lesssim k \left(\log \ell + \log \|\theta\|_{\mathcal{H}_1} + \log \frac{1}{\epsilon} \right).$$

We then naturally provide a description of the situation when f is known and parametrized by an infinite sequence $\theta \in \ell^2(\mathbb{Z})$. According to the previous computations, and using a truncation argument at frequency $\ell_{\epsilon} = \epsilon^{-1/s}$ in the Sobolev space \mathcal{H}_s , one can show the following result.

COROLLARY 1. *Assume $f \in \mathcal{H}_s$ known for $s \geq 1$ ($\theta := \theta(f)$ such that $\sum_{j \in \mathbb{Z}} |\theta_j|^2 |j|^{2s} < +\infty$), using the same set \mathcal{A}_{θ} as in Proposition 3.1 with $\ell_{\epsilon} = \epsilon^{-1/s}$, then*

$$H_{\square}(\epsilon, \mathcal{A}_{\theta}, d_H) \lesssim \frac{s+1}{s} \log \frac{1}{\epsilon} + \log \|\theta\|_{\mathcal{H}_1}.$$

Similarly, one also has

$$H_{\square}(\epsilon, \mathcal{M}_{\theta}^k, d_H) \lesssim k \left(\frac{s+1}{s} \log \frac{1}{\epsilon} + \log \|\theta\|_{\mathcal{H}_1} \right).$$

The next step is to consider a continuous mixture for g , which is the more natural case. For f known, let

$$\mathcal{P}_f := \{\mathbb{P}_{f,g} \mid g \in \mathfrak{M}([0, 1])\}.$$

Once again, we will only consider functions f with null Fourier coefficients of order higher than ℓ_ϵ . For sake of simplicity, we will omit the dependence on ϵ with the notation ℓ .

It would be quite tempting to use the results of [GvdV01] to bound the bracketing entropy of \mathcal{P}_f , but indeed as pointed by [MM11] applying directly the bounds obtained in Lemma 3.1 and Lemma 3.2 of [GvdV01] to our setting yields a too weak result: the size of the upper bound on $H_{[]}(\epsilon, \mathcal{P}_f, d_H)$ will have a too strong dependency on ℓ . By the way, we have to carefully adapt the approach of [GvdV01] to obtain a sufficiently sharp upper bound of the entropy of \mathcal{P}_f . Such bound is given in the next result, in which we provide a majorization of the entropy with respect to the Total Variation distance which is easier to handle here. Note that all the previous results are still true if we use d_H instead of d_{TV} since (A.2) also permits to retrieve entropy bounds for d_H from entropy bounds for d_{TV} .

PROPOSITION 3.3. *Let $\epsilon > 0$ and $s > 0$, if $\log \frac{1}{\epsilon} \lesssim \ell$ and $f \in \mathcal{H}^\ell$ is such that $\|\theta\|^2 \lesssim 2\ell + 1$, then*

$$\log N(\epsilon, \mathcal{P}_f, d_{TV}) \lesssim \ell^2 \left(\log \frac{1}{\epsilon} + \log \|\theta\|_{\mathcal{H}_1} \right).$$

If furthermore $w \lesssim \sqrt{2\ell + 1}$ then

$$\sup_{f \in \mathcal{H}^\ell : \|\theta(f)\| \leq w} \log N(\epsilon, \mathcal{P}_f, d_{TV}) \lesssim \ell^2 \left(\log \frac{1}{\epsilon} + \log \ell \right).$$

The second inequality opens the way for the case of unknown f given below. It is possible since in the first inequality we have carefully expressed the dependency on f and ℓ .

The method to build an ϵ -covering of \mathcal{P}_f follows two natural steps:

- approximate any mixture g by a finite one \tilde{g} such that

$$d_{TV}(\mathbb{P}_{\theta,g}, \mathbb{P}_{\theta,\tilde{g}}) \leq \epsilon/2,$$

with a number of components of the finite mixture \tilde{g} uniformly bounded in g (depending on f and ϵ);

- use Proposition 3.2 for the finite mixture to well approximate $\mathbb{P}_{\theta,\tilde{g}}$.

The proof itself is delayed to the Appendix.

3.1.2. *Case of unknown f .* We now describe the picture when f is unknown, which is the main objective of this paper. We assume that f belongs to \mathcal{H}_s . In order to bound the bracketing entropy, we define a sieve over \mathcal{H}_s which depends on a frequency cut-off ℓ and a size parameter w . We then get

$$\mathcal{P}_{\ell,w} := \left\{ \mathbb{P}_{f,g} \mid f \in \mathcal{H}_s^\ell, \|\theta(f)\| \leq w, g \in \mathfrak{M}([0,1]) \right\}.$$

THEOREM 3.1. *Let be given $\epsilon > 0$ small enough, and assume that ℓ_ϵ and w_ϵ are such that $\log \frac{1}{\epsilon} \lesssim \ell_\epsilon$ and $w_\epsilon \lesssim \sqrt{\ell_\epsilon}$, then*

$$\log N(\epsilon, \mathcal{P}_{\ell_\epsilon, w_\epsilon}, d_{TV}) \lesssim \ell_\epsilon^2 \left(\log \frac{1}{\epsilon} + \log \ell_\epsilon \right).$$

The proof of Theorem 3.1 is based on two simple results. The first one is the Girsanov formula obtained by [BG10] in appendix A.2.2 (in the case of known g): it can be extended to the situation of unknown g and complex trajectories as in (2.1), which leads to

$$(3.1) \quad \frac{d\mathbb{P}_{f,g}}{d\mathbb{P}_{f^0,g^0}}(Y) = \frac{\int_0^1 \exp(2\Re\langle f^{-\alpha_1}, dY \rangle - \|f^{-\alpha_1}\|^2) dg(\alpha_1)}{\int_0^1 \exp(2\Re\langle f^{0,-\alpha_2}, dY \rangle - \|f^{0,-\alpha_2}\|^2) dg^0(\alpha_2)},$$

for any measurable trajectory Y .

The second result is given in the following lemma.

LEMMA 3.1. *Let f and \tilde{f} be any functions in $L^2_{\mathbb{C}}([0,1])$, g be any shift distribution in $\mathfrak{M}([0,1])$, then*

$$d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{\tilde{f},g}) \leq \frac{\|f - \tilde{f}\|}{\sqrt{2}}.$$

PROOF OF THEOREM 3.1. The idea of the demonstration is to build a ϵ -covering of $\mathcal{P}_{\ell,w}$ with $\epsilon/2$ -coverings for f and g . First, let $\mathbb{P}_{f,g}$ and $\mathbb{P}_{\tilde{f},\tilde{g}}$ two elements of $\mathcal{P}_{\ell,w}$ and remark that by the triangle inequality

$$d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{\tilde{f},\tilde{g}}) \leq d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{\tilde{f},g}) + d_{TV}(\mathbb{P}_{\tilde{f},g}, \mathbb{P}_{\tilde{f},\tilde{g}}).$$

We will look for a covering method that will use the inequality above and a tensorial argument, it requires to bound both terms. The majorization of the first one comes from Lemma 3.1. The second term is handled uniformly in \tilde{f} by Proposition 3.3.

Now, we build $\epsilon/2$ -coverings of $\mathbb{P}_{f,g}$ for fixed g from an $\epsilon/\sqrt{2}$ -covering of f for the L^2 -norm:

$$\log N\left(\epsilon/\sqrt{2}, \left\{f \in \mathcal{H}_s^{\ell_\epsilon}, \|\theta(f)\| \leq w_\epsilon\right\}, \|\cdot\|\right) \lesssim \ell_\epsilon \log \frac{w_\epsilon}{\epsilon} = o\left(\ell_\epsilon^2 \log \frac{1}{\epsilon}\right).$$

□

According to inequality (A.2) and since $\log \frac{1}{\epsilon^2} \lesssim \log \frac{1}{\epsilon}$, we can easily deduce the next corollary.

COROLLARY 2. *Let be given $\epsilon > 0$ small enough, and assume $\log \frac{1}{\epsilon} \lesssim \ell_\epsilon$ and $w_\epsilon \leq \sqrt{2\ell_\epsilon + 1}$, then*

$$\log N(\epsilon, \mathcal{P}_{\ell_\epsilon, w_\epsilon}, d_H) \lesssim \ell_\epsilon^2 \left(\log \frac{1}{\epsilon} + \log \ell_\epsilon \right).$$

REMARK 3.1. *i) Even if the model studied here is a very special case of Gaussian mixture models, one may think that such kind of results may help the analysis of more general mixture cases within a growing dimension setting.*

ii) In our case, we will use a much higher choice of ℓ_ϵ than $\log \frac{1}{\epsilon}$. This choice will be fixed in section 4.1.

3.2. Link between Kullback-Leibler and Hellinger neighbourhoods. We first recall a useful result of Wong & Shen given as Theorem 5 in [WS95]. It enables to handle Hellinger neighbourhoods instead of $\mathcal{V}_{\epsilon_n}(\mathbb{P}_{f^0, g^0}, d_{KL})$, which is generally easier for mixture models.

THEOREM 3.2 (Wong & Shen). *Let μ and ν be two measures such that μ is a.c. with respect to ν with a density $q = d\mu/d\nu$. Assume that $d_H(\mu, \nu)^2 = \int [\sqrt{q} - 1]^2 d\nu \leq \epsilon^2$ and that there exists $\delta \in (0, 1]$ such that*

$$(3.2) \quad M_\delta^2 := \int_{q \geq \epsilon^{1/\delta}} q^{\delta+1} d\nu < \infty.$$

Then, for ϵ small enough, there exists a universal constant C large enough such that

$$d_{KL}(\mu, \nu) = \int q \log q d\nu \leq C \log(M_\delta) \epsilon^2 \log \frac{1}{\epsilon},$$

and

$$V(\mu, \nu) \leq \int q \log^2 q d\nu \leq C \log(M_\delta)^2 \epsilon^2 \left[\log \frac{1}{\epsilon} \right]^2.$$

Hence, Hellinger neighbourhoods are almost Kullback-Leibler ones (up to some logarithm terms) provided that a sufficiently large moment exists for q ($q \log q$ is killed by $q^{1+\delta}$ for large values of q and a second order expansion of $q \log q - q + 1$ around 1 yields a term similar to $[\sqrt{q} - 1]^2$). Next proposition shows that condition (3.2) is satisfied in our SIM.

PROPOSITION 3.4. *For any $\mathbb{P}_{f^0, g^0} \in \mathcal{P}$, and for any $f \in \mathcal{H}_s$ such that $\|f\| \leq 2\|f^0\|$, and any $g \in \mathfrak{M}([0, 1])$, define $q = \frac{d\mathbb{P}_{f^0, g^0}}{d\mathbb{P}_{f, g}}$. There exists $\delta \in (0, 1]$ such that the constant defined in equation (3.2) M_δ^2 is uniformly bounded with respect to f .*

3.3. *Hellinger neighbourhoods.* Proposition 3.4 will allow to use Theorem 3.2, thus we now aim to find a lower bound on Hellinger neighbourhood of \mathbb{P}_{f^0, g^0} . Consider a frequency cut-off ℓ_n that will be fixed later. For any $f \in \mathcal{H}_s^{\ell_n}$ and $g \in \mathfrak{M}([0, 1])$, remind that we denote $\theta := \theta(f)$ as well as $\theta^0 = \theta(f^0)$. We define $f_{\ell_n}^0$ the L^2 projection of f^0 on the subspace $\mathcal{H}_s^{\ell_n}$.

For sake of simplicity, $\mathbb{E}_0 F(Y)$ will refer to the expectation of a function F of the trajectory Y when Y follows \mathbb{P}_{f^0, g^0} . The triangle inequality applied to the Hellinger distance shows that

$$d_H(\mathbb{P}_{f^0, g^0}, \mathbb{P}_{f, g}) \leq \overbrace{d_H(\mathbb{P}_{f^0, g^0}, \mathbb{P}_{f_{\ell_n}^0, g^0})}^{(E_1)} + \overbrace{d_H(\mathbb{P}_{f_{\ell_n}^0, g^0}, \mathbb{P}_{f_{\ell_n}^0, g})}^{(E_2)} + \overbrace{d_H(\mathbb{P}_{f_{\ell_n}^0, g}, \mathbb{P}_{f, g})}^{(E_3)}.$$

In the sequel, we will provide sufficiently sharp upper bound on (E_1) , (E_2) , (E_3) so that we will be able to find a suitable lower bound of the prior mass of Hellinger neighbourhoods.

Upper bound of (E_1) . We first bound (E_1) using $d_H^2 \leq d_{KL}$ with the Girsanov formula (3.1)

$$\begin{aligned} (E_1) &\leq \sqrt{d_{KL}(\mathbb{P}_{f^0, g^0}, \mathbb{P}_{f_{\ell_n}^0, g^0})} \\ &= \left(\mathbb{E}_0 \left[-\log \frac{\int_0^1 \exp \left(2\Re \langle f_{\ell_n}^{0, -\alpha}, dY \rangle - \|f_{\ell_n}^0\|^2 \right) dg^0(\alpha)}{\int_0^1 \exp \left(2\Re \langle f^{0, -\alpha}, dY \rangle - \|f^0\|^2 \right) dg^0(\alpha)} \right] \right)^{1/2} \\ &:= (\tilde{E}_1) \end{aligned}$$

We now obtain the upper bound of (E_1) according to the next proposition.

PROPOSITION 3.5. *Assume that $Y \sim \mathbb{P}_{f^0, g^0}$ and $f^0 \in \mathcal{H}_s$, then*

$$(E_1) \leq (\tilde{E}_1) \leq \sqrt{2}\|f^0 - f_{\ell_n}^0\| \leq \sqrt{2}\|f^0\|_{\mathcal{H}_1} \ell_n^{-s}.$$

Upper bound of (E_3) . We will be interested in the Hellinger distance when $f_{\ell_n}^0$ is close to f , and the dimension ℓ_n grows up to $+\infty$ (the mixture law on $[0, 1]$ is the same for the two laws). The important fact will be its exclusive dependence with respect to the L^2 distance between $f_{\ell_n}^0$ and f . This

upper bound is given in the next proposition, whose proof is immediate from Lemma 3.1 and equation (A.2).

PROPOSITION 3.6. *Assume that $f \in \mathcal{H}_s^{\ell_n}$ and $g \in \mathfrak{M}([0, 1])$, then*

$$d_H(\mathbb{P}_{f_{\ell_n}^0, g}, \mathbb{P}_{f, g}) \leq 2^{1/4} \sqrt{\|f - f_{\ell_n}^0\|}.$$

Upper bound for (E_2) . This term is clearly the more difficult to handle. We will obtain a convenient result using some elements obtained in Proposition 3.3. For a given $\epsilon_n > 0$, $\ell_n, f_{\ell_n}^0 \in \mathcal{H}_s^{\ell_n}$ and $g^0 \in \mathfrak{M}([0, 1])$, we know that one may find a mixture model \tilde{g} such that $d_H(\mathbb{P}_{f_{\ell_n}^0, g^0}, \mathbb{P}_{f_{\ell_n}^0, \tilde{g}}) < \epsilon_n$ and \tilde{g} has $C\ell_n^2$ points of support in $[0, 1]$ as soon as ϵ_n is small enough and $\log \frac{1}{\epsilon_n} \lesssim \ell_n$ (the condition $\|f_{\ell_n}^0\|^2 \leq 2\ell_n + 1$ is immediate since f^0 does not depend on n). The next step is to control the Hellinger distance $d_H(\mathbb{P}_{f_{\ell_n}^0, g}, \mathbb{P}_{f_{\ell_n}^0, \tilde{g}})$ for $g \in \mathfrak{M}([0, 1])$, and this can be done thanks to an adaptation in dimension $2\ell_n + 1$ of Lemma 5.1 of [GvdV01].

LEMMA 3.2. *Let be given \tilde{g} a discrete mixture law whose support is of cardinal J whose support points $(\varphi_j)_{j=1\dots J}$ are such that $\tilde{g}(\varphi_j) = p_j$ and η -separated, i.e. $|\varphi_j - \varphi_i| \geq \eta, \forall i \neq j$, then $\forall \tilde{g} \in \mathfrak{M}([0, 1])$*

$$d_H^2(\mathbb{P}_{f_{\ell_n}^0, \tilde{g}}, \mathbb{P}_{f_{\ell_n}^0, g}) \leq \sqrt{\frac{\pi}{2}} \|f_{\ell_n}^0\|_{\mathcal{H}_1} \eta + 2 \sum_{j=1}^J |\tilde{g}([\varphi_j - \eta/2, \varphi_j + \eta/2]) - \tilde{g}(\varphi_j)|.$$

In [BG13], we will show that it is possible to obtain a more general upper bound for the Hellinger distance between $\mathbb{P}_{f_{\ell_n}^0, \tilde{g}}$ and $\mathbb{P}_{f_{\ell_n}^0, g}$ which implies the Wasserstein distance $W_1(g, \tilde{g})$ between g and \tilde{g} , but such upper bound is a little bit less powerful than the one given by the former lemma. Note that Lemma 3.2 needs a discrete mixture with η -separated support points. The following result permits to obtain such a mixture.

PROPOSITION 3.7. *Assume that $f^0 \in \mathcal{H}_s$ for $s \geq 1$, $g^0 \in \mathfrak{M}([0, 1])$, and $\log \frac{1}{\epsilon_n} \lesssim \ell_n$. For any $\eta_n \leq \epsilon_n^2$, there exists a discrete distribution \tilde{g} with in its support at most $J_n \lesssim \ell_n^2$ points denoted $(\psi_j)_{j=1\dots J_n}$, such that these points are η_n -separated, and*

$$d_H(\mathbb{P}_{f_{\ell_n}^0, g^0}, \mathbb{P}_{f_{\ell_n}^0, \tilde{g}}) \leq \left(1 + (8\pi)^{1/4} \|f_{\ell_n}^0\|_{\mathcal{H}_1}^{1/2}\right) \epsilon_n.$$

Furthermore, for any $g \in \mathfrak{M}([0, 1])$,

$$d_H(\mathbb{P}_{f_{\ell_n}^0, g^0}, \mathbb{P}_{f_{\ell_n}^0, g}) \leq \left(1 + (8\pi)^{1/4} \|f^0\|_{\mathcal{H}_1}^{1/2}\right) \epsilon_n \\ + \sqrt{\frac{\pi}{2} \|f^0\|_{\mathcal{H}_1} \eta_n + 2 \sum_{j=1}^{J_n} |g(\psi_j - \eta_n/2, \psi_j + \eta_n/2) - \tilde{g}(\psi_j)|}.$$

In [BG13] we obtain a more general upper bound for (E_2) , based on the Wasserstein distance. We could use it to retrieve Proposition 3.7, but it also leads to Hellinger neighbourhoods described in terms of the Total Variation distance from g to g^0 . This last distance is adapted to smooth densities g but not to the ones considered here, when the prior distribution for g is a Dirichlet process.

Description of a Hellinger neighbourhood. We can now gather the upper bounds of (E_1) , (E_2) , and (E_3) to get the following result.

PROPOSITION 3.8. *Assume that $f^0 \in \mathcal{H}_s$ for $s \geq 1$ and $g^0 \in \mathfrak{M}([0, 1])$. Choose the threshold such as $\epsilon_n^{-1/s} \lesssim \ell_n \lesssim \epsilon_n^{-1/s}$ and $\eta_n := \epsilon_n^2$, and consider the finite mixture \tilde{g} provided by Proposition 3.7. Define*

$$\mathcal{G}_{\epsilon_n} := \left\{ g \in \mathfrak{M}([0, 1]) : \sum_{j=1}^{J_n} |g(\psi_j - \eta_n/2, \psi_j + \eta_n/2) - \tilde{g}(\psi_j)| \leq \epsilon_n^2 \right\}, \\ \mathcal{F}_{\epsilon_n} := \left\{ f \in \mathcal{H}_s^{\ell_n} : \|f - f_{\ell_n}^0\| \leq \epsilon_n^2 \right\}.$$

Then, there exists a constant C_0 depending only on $\|f^0\|_{\mathcal{H}_1}$ such that for any $g \in \mathcal{G}_{\epsilon_n}$ and $f \in \mathcal{F}_{\epsilon_n}$,

$$d_H(\mathbb{P}_{f^0, g^0}, \mathbb{P}_{f, g}) \leq C_0 \epsilon_n.$$

4. Proof of Theorem 2.2. We will prove this result using the "tool-box" provided by Theorem 2.1. We thus check its applicability and consider each of its hypotheses.

4.1. *Checking the conditions of Theorem 2.1.* We first prove the minoration for the lower bound (2.5), necessary to apply Theorem 2.1.

PROPOSITION 4.1. *Assume that $f^0 \in \mathcal{H}_s$ for $s \geq 1$ and $g^0 \in \mathfrak{M}([0, 1])$. For any sequence $(\epsilon_n)_{n \in \mathbb{N}}$ which converges to 0 as $n \rightarrow +\infty$, and for the prior defined in paragraph 2.3, there exists a constant $c > 0$ such that*

$$\Pi_n(\mathbb{P}_{f, g} \in \mathcal{P} : d_{KL}(\mathbb{P}_{f, g}, \mathbb{P}_{f^0, g^0}) \leq \epsilon_n^2, V(\mathbb{P}_{f, g}, \mathbb{P}_{f^0, g^0}) \leq \epsilon_n^2) \geq h_n,$$

where

$$h_n := e^{-(c+o(1)) \left[\epsilon_n^{-2/s} (\log(1/\epsilon_n))^{\rho+2/s} \vee \xi_n^{-2} \right]}.$$

Proposition 4.1 relies on Theorem 3.2, which permits to use Hellinger neighbourhoods instead of $\mathcal{V}_{\epsilon_n}(\mathbb{P}_{f^0, g^0}, d_{KL})$, and on Proposition 3.8, which describes suitable Hellinger neighbourhoods. To control their prior mass, we remind the following useful result appeared as Lemma 6.1 of [GGvdV00]. This enables to find a lower bound of ℓ_1 -ball of radius r under Dirichlet prior.

LEMMA 4.1 ([GGvdV00]). *Let $r > 0$ and (X_1, \dots, X_N) be distributed according to the Dirichlet distribution on the ℓ_1 simplex of dimension $N - 1$ with parameters $(m, \alpha_1, \dots, \alpha_N)$. Assume that $\sum_j \alpha_j = m$ and $Ar^b \leq \alpha_j \leq 1$ for some constants A and b . Let (x_1, \dots, x_N) be any points on the N simplex, there exists c and C that only depend on A and b such that if $r \leq 1/N$*

$$\Pr \left(\sum_{j=1}^N |X_j - x_j| \leq 2r \right) \geq C \exp \left(-cN \log \frac{1}{r} \right)$$

In the proof of Proposition 4.1 (delayed to the Appendix), one can see that we could obtain a suitable lower bound as soon as $\lambda(\ell_n) \geq e^{-c\ell_n^2 \log \ell_n}$ for a constant c . Of course, a distribution λ with some heavier tail would also suit here. However, such a heavier tail is not suitable for the control of the term $\Pi_n(\mathcal{P} \setminus \mathcal{P}_n)$ which is detailed in the next proposition.

PROPOSITION 4.2. *For any sequences $k_n \mapsto +\infty$ and $\epsilon_n \mapsto 0$ as $n \mapsto +\infty$, define $w_n^2 = 4k_n + 2$, then there exists a constant c such that*

$$\Pi_n(\mathcal{P} \setminus \mathcal{P}_{k_n, w_n}) \leq e^{-c[k_n^2 \log^\rho(k_n) \wedge k_n \xi_n^{-2}]},$$

and

$$\log D(\epsilon_n, \mathcal{P}_{k_n, w_n}, d_H) \lesssim k_n^2 \left[\log k_n + \log \frac{1}{\epsilon_n} \right].$$

We are now able to conclude the proof of the posterior consistency.

PROOF OF THEOREM (2.2). Take $\epsilon_n := n^{-\alpha}(\log n)^\kappa$ and $k_n := n^\beta(\log n)^\gamma$. From our definition (2.6), we have also $\xi_n^{-2} = n^{\mu_s}(\log n)^\zeta$, and we look for admissible values of α , β , κ , γ , μ_s , and ζ in order to satisfy (2.3), (2.4) and (2.5).

Proposition (4.1) imposes that in order to satisfy (2.5), we could check that

$$\epsilon_n^{-2/s} \left(\log \frac{1}{\epsilon_n} \right)^{\rho+2/s} \vee n^{\mu_s} (\log n)^\zeta \ll n \epsilon_n^2 = n^{1-2\alpha} (\log n)^{2\kappa}.$$

This is true as soon as ϵ_n satisfies

$$\alpha \leq \frac{s}{2s+2} \quad \text{and} \quad \kappa > (\rho s + 2)/(2s + 2).$$

Moreover, we obtain the first condition on μ_s : $\mu_s \leq 1 - 2\alpha$, and if $\mu_s = 1 - 2\alpha$ then $\zeta < 2\kappa$.

Now, Proposition (4.2) shows that (2.3) is fulfilled provided that

$$(4.1) \quad k_n^2 \left[\log k_n + \log \frac{1}{\epsilon_n} \right] \lesssim n \epsilon_n^2 = n^{1-2\alpha} (\log n)^{2\kappa}.$$

This condition is satisfied when $2\beta \leq 1 - 2\alpha$ and $2\gamma + 1 \leq 2\kappa$. At last, Proposition (4.2) again ensures that (2.4) is true as soon as

$$k_n^2 \log^\rho k_n \wedge k_n n^{\mu_s} \gtrsim n \epsilon_n^2$$

and we deduce from (4.1) that

$$2\beta = 1 - 2\alpha \quad \text{and} \quad -\rho/2 + \kappa \leq \gamma \leq -1/2 + \kappa.$$

Moreover, we also see that $\beta + \mu_s \geq 1 - 2\alpha$, hence $\mu_s \geq 1/2 - \alpha$, and if $\mu_s = 1/2 - \alpha$ then $\gamma + \zeta \geq 2\kappa$; the former condition on μ_s yields $\mu_s \geq 1/2 - \alpha \geq \frac{1}{2s+2}$ (which naturally drives us to set $\mu_s = 1/4$ (case $s = 1$) for adaptive prior).

We split the proof according to the adaptive or non adaptive case.

Adaptive prior. We first set μ independent of s and equal to $1/4$. For any $s \in [1, 3]$, we see that $\alpha(s) = s/(2s + 2)$ is the admissible largest value of α and $\alpha(s) = 3/8 < s/(2s + 2)$ as soon as $s > 3$. The corresponding value of β is $1/(2s + 2)$ when $s \in [1, 3]$ and $\beta = 1/8$ otherwise. Any choice of $\zeta \in [3/2, 2)$ permits to deal with the conditions on ζ that appears when $s = 1$ or $s \geq 3$. The other values of γ and κ may be determined with respect to ρ . For instance, if we choose $\rho \in (1, 2)$, we can take $\kappa = 1$ and $\gamma = 1/2$.

Non adaptive prior. The non adaptive case is much more simpler since it is sufficient to fix

$$\mu_s = 1 - 2\alpha = 2/(2s + 2)$$

and $\zeta = 0$ to obtain suitable calibrations for α, β, κ and γ . This achieves the proof. \square

5. Concluding remarks. In this paper, we exhibit a suitable prior which enable to obtain a contraction rate of the posterior distribution near the true underlying distribution \mathbb{P}_{f^0, g^0} . Moreover, this rate is polynomial with the number n of observations, even if our SIM is an inverse problem with unknown operator of translation which depends on g . From a technical point of view, the keystones of such results are the tight link between the white noise model and the Fourier expansion as well as the smoothness of Gaussian law which permits to obtain an efficient covering strategy.

A natural problem would study of the behaviour of the posterior distribution regarding the functional objects shape f^0 and mixture law g^0 . This question is tackled in [BG13] where we establish a contraction of the posterior distribution around f^0 and g^0 up to identifiability conditions.

Another interesting extension would consider the SIM with a noise level σ depending on n in the Bayesian framework. This asymptotic setting is linked to the work of [BG12] in which their J curves are sampled at the n points of a discrete design in $[0, 1]$.

At last, an open and challenging question concerns the research of stochastic algorithm to approach the posterior distribution in our non parametric Shape Invariant Model. One may think of an adaptation of the SA-EM strategy proposed in [AKT10] even if this approach is at the moment valid only in a parametric setting.

APPENDIX A: TOPOLOGY ON PROBABILITY SPACE

Probability distances. We study consistency using standard distance over probability measures. If P and Q are two probability measures over a set X , absolutely continuous with respect to a reference measure λ , d_H refers to the Hellinger distance defined as

$$d_H(P, Q) := \sqrt{\int_X \left[\sqrt{\frac{dP}{d\lambda}} - \sqrt{\frac{dQ}{d\lambda}} \right]^2 d\lambda}.$$

Note that d_H does not depend on the choice of the dominating measure λ , and that the definition can be extended to any finite measures P and Q in a straightforward way.

When needed, we use the Total Variation distance between two probability measures P and Q . If \mathcal{B} is the σ -algebra of measurable sets with the reference measure λ , this distance is given by

$$d_{TV}(P, Q) := \sup_{A \in \mathcal{B}} |P(A) - Q(A)| = \frac{1}{2} \int_X \left| \frac{dP}{d\lambda} - \frac{dQ}{d\lambda} \right| d\lambda.$$

At last, we recall the definition of the Kullback-Leibler divergence (entropy) between P and Q since it is sometimes be used in the work:

$$d_{KL}(P, Q) := \int_X -\log \frac{dQ}{dP} dP.$$

In the sequel, we shall also use $V(P, Q)$ defined as a second order moment associated to the Kullback-Leibler divergence

$$V(P, Q) := \int_X \left(\log \frac{dQ}{dP} \right)^2 dP.$$

It may be reminded the classical Pinsker's inequality

$$(A.1) \quad \sqrt{\frac{1}{2} d_{KL}(P, Q)} \geq d_{TV}(P, Q),$$

as well as

$$(A.2) \quad \frac{1}{2} d_H(P, Q)^2 \leq d_{TV}(P, Q) \leq d_H(P, Q).$$

Model Complexity. To obtain the posterior consistency and convergence rate, we shall use results given by Theorem 2.1 of [GGvdV00] which is stated below. This theorem exploits the notion of complexity of the studied model, and this complexity is traduced according to packing or covering numbers. For any set of probability measures \mathcal{P} endowed with a metric d , $D(\epsilon, \mathcal{P}, d)$ refers to the ϵ -packing number (the maximum number of points in \mathcal{P} such that the minimal distance between each pair is larger than ϵ). The ϵ -covering number $N(\epsilon, \mathcal{P}, d)$ is the minimum number of balls of radius ϵ needed to cover \mathcal{P} . These two numbers are linked through the following inequality

$$N(\epsilon, \mathcal{P}, d) \leq D(\epsilon, \mathcal{P}, d) \leq N(\epsilon/2, \mathcal{P}, d).$$

At last, for d a metric on finite measures, an ϵ -bracket is a set of the form

$$[L, U] := \left\{ P \text{ s.t. } \frac{dL}{d\lambda} \leq \frac{dP}{d\lambda} \leq \frac{dU}{d\lambda} \right\},$$

for L and U two finite measures such that $d(L, U) \leq \epsilon$ and λ any dominating measure. The ϵ -bracketing number $N_{[]}(\epsilon, \mathcal{P}, d)$ is the minimal number of ϵ -brackets needed to cover \mathcal{P} . Note that $N_{[]}(\epsilon, \mathcal{P}, d_H)$ is an upper bound of the $(\epsilon/2)$ -covering number $N(\epsilon/2, \mathcal{P}, d_H)$. The bracketing entropy is then defined by $H_{[]}(\epsilon, \mathcal{P}, d) := \log N_{[]}(\epsilon, \mathcal{P}, d)$.

APPENDIX B: TOOLS FOR THE PROOF OF THEOREM 2.2

B.1. Entropy estimates.

PROOF OF PROPOSITION 3.1. The proof is similar to Lemma 1 of [GW00], we set $p = 2\ell + 1$ and for any $\epsilon > 0$, we are going to build an explicit bracketing of \mathcal{A}_θ and then bound $N_{[]}(\epsilon, \mathcal{A}_\theta, d_H)$. For an integer K which will be chosen in the sequel, we define $[\varphi_-^i, \varphi_+^i]$ of size $\Delta_\varphi = 1/K$, with $\varphi_-^i = (i-1)\Delta_\varphi$ and $\varphi_+^i = i\Delta_\varphi$. For any $\delta > 0$, we consider the lower and upper brackets

$$l_i := (1 + \delta)^{-1} \gamma_{\theta \bullet \varphi_-^i, (1+\delta)^{-\alpha} Id} \quad \text{and} \quad u_i := (1 + \delta) \gamma_{\theta \bullet \varphi_+^i, (1+\delta)^\alpha Id}.$$

We are looking for some admissible values of α , δ , and K such that the set $([l_i, u_i])_{i=1 \dots K}$ is an ϵ -bracket of \mathcal{A}_θ for d_H . Of course, for all $\varphi \in [\varphi_-^i, \varphi_+^i]$, $l_i \leq \gamma_{\theta \bullet \varphi, Id}(\cdot) \leq u_i$ should hold, but we can check that $\forall x \in \mathbb{C}$,

$$\frac{l_i(x)}{\gamma_{\theta \bullet \varphi, Id}(x)} \leq \frac{1}{1 + \delta} \frac{1}{(1 + \delta)^{-p\alpha}} e^{\frac{\|\theta \bullet \varphi - \theta \bullet \varphi_-^i\|^2}{1 - (1 + \delta)^{-\alpha}}} \leq (1 + \delta)^{p\alpha - 1} e^{\frac{4\pi^2 \Delta_\varphi^2 \|\theta\|_{\mathcal{H}_1}^2}{1 - (1 + \delta)^{-\alpha}}}.$$

Hence, we must have $\alpha \leq 1/p$, and we must also satisfy

$$|\Delta_\varphi|^2 \leq \frac{1 - p\alpha}{4\pi^2 \|\theta\|_{\mathcal{H}_1}^2} (1 - (1 + \delta)^{-\alpha}) \log(1 + \delta) = \frac{\alpha(1 - p\alpha)\delta^2}{4\pi^2 \|\theta\|_{\mathcal{H}_1}^2} (1 + o(1)),$$

where $o(1)$ does not depend on p and goes to zero as $\delta \rightarrow 0$ uniformly in α in any positive neighbourhood of zero. In a same way considering $\gamma_{\theta \bullet \varphi, Id} u_i^{-1}$, we obtain

$$\forall x \in \mathbb{C} \quad \frac{\gamma_{\theta \bullet \varphi, Id}(x)}{u_i(x)} \leq (1 + \delta)^{\alpha p - 1} e^{\frac{4\pi^2 \Delta_\varphi^2 \|\theta\|_{\mathcal{H}_1}^2}{(1 + \delta)^{\alpha - 1}}},$$

and the same conditions arise. In order to minimize the cardinal of the bracketing, Δ_φ must be as large as possible, we then maximize $\alpha(1 - p\alpha)$ and choose $\alpha = (2p)^{-1}$.

We must now check that $d_H(l_i, u_i) \leq \epsilon$. Rapid computations show that

$$d_H(l_i, u_i)^2 = \delta^2 + d_H(\gamma_{\theta \bullet \varphi_-^i, (1+\delta)^{-\alpha} Id}(\cdot), \gamma_{\theta \bullet \varphi_+^i, (1+\delta)^\alpha Id}(\cdot))^2.$$

Using standard formula on Hellinger distance for multivariate gaussian laws, we obtain

$$\begin{aligned} d_H(l_i, u_i)^2 &= \delta^2 + 2 \left[1 - \frac{2^p}{((1 + \delta)^\alpha + (1 + \delta)^{-\alpha})^p} \right] \\ &= \delta^2 + 2 \left[1 - \frac{2^p \sqrt{1 + \delta}}{(1 + (1 + \delta)^{1/p})^p} \right]. \end{aligned}$$

One can easily check that, whatever $p \geq 1$, $(1 + (1 + \delta)^{1/p})^p \leq 2^p e^{\delta/2}$, which yields

$$d_H(l_i, u_i)^2 \leq \frac{3}{2}\delta^2 + o(\delta^2) \leq 2\delta^2$$

for δ small enough. An admissible choice of δ should be $\delta = \epsilon/\sqrt{2}$, which insures $d_H(l_i, u_i) \leq \epsilon$. We then obtain

$$\Delta_\varphi^2 \leq \frac{\delta^2 + o(\delta^2)}{16\pi^2 p \|\theta\|_{\mathcal{H}_1}^2} = \frac{\epsilon^2 + o(\epsilon^2)}{32\pi^2 p \|\theta\|_{\mathcal{H}_1}^2},$$

where $o(\epsilon^2)$ does not depend on p . The number of brackets is now $K = \Delta_\varphi^{-1}$, this ends the proof of the proposition. \square

PROOF OF PROPOSITION 3.3. We first fix the notation $p = 2\ell + 1$ which refers to the dimension of the multivariate mixture. For any $R > 0$ which will be chosen later, \mathcal{E}_R is the ball of in \mathbb{C}^p of radius R . For sake of simplicity, we will sometimes omit the dependence on ϵ with the notation p . According to the hypotheses in Proposition 3.3, there exists an absolute constant a such that $\|\theta\| \leq w \leq a\sqrt{p}$. We first write

$$d_{TV}(\mathbb{P}_{\theta,g}, \mathbb{P}_{\theta,\tilde{g}}) \leq \underbrace{\frac{1}{2} \int_{\mathcal{E}_R^c} |d\mathbb{P}_{\theta,g} - d\mathbb{P}_{\theta,\tilde{g}}|(z)}_{:= (A)} + \underbrace{\frac{1}{2} \int_{\mathcal{E}_R} |d\mathbb{P}_{\theta,g} - d\mathbb{P}_{\theta,\tilde{g}}|(z)}_{:= (B)}.$$

Let ν be a measure on $[0, 1]$ that dominates both g and \tilde{g} .

Term (A). We will pick R such that (A) is smaller than $\epsilon/2$, first set $R^2 > (1+a)^2 p \geq a^{-2}(1+a)^2 \|\theta\|^2$ and with this choice,

$$\forall \varphi \in [0, 1] \quad \forall z \in \mathcal{E}_R^c \quad \|z - \theta \bullet \varphi\| > \|z\|/(1+a).$$

This simply implies that,

$$\begin{aligned} (A) &\leq \pi^{-p} \int_{\mathcal{E}_R^c} \int_0^1 e^{-\frac{\|z\|^2}{(1+a)^2}} \left| \frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right| d\nu(\varphi) dz \\ &\leq 2(1+a)^{2p} \mathbb{P} \left(\chi_{2p}^2 \geq \frac{2R^2}{(1+a)^2} \right). \end{aligned}$$

To deal with we last term we use a concentration of chi-square statistics inequality (see Lemma 1 of [IL06]): for any $k \geq 1$ and $c > 0$,

$$(B.1) \quad \mathbb{P}(\chi_k^2 \geq (1+c)k) \leq \frac{1}{c\sqrt{2\pi}} e^{-\frac{k}{2}[c - \log(1+c)] - \frac{1}{2} \log k}.$$

Therefore, writing $R^2 = (1+a)^2(1+c)p$ for $c > 0$, one gets

$$(A) \leq \frac{1}{c\sqrt{\pi}} e^{-p[c - \log(1+c) - 2\log(1+a)] - \frac{1}{2}\log p}$$

and this term is smaller than $\epsilon/2$ if we pick c large enough, since $\log \frac{1}{\epsilon} \lesssim p$.

Term (B). We then consider (B), following the strategy of [GvdV01] which exploits the smoothness of Gaussian densities. We will exhibit a discrete mixture law which will be close to $\mathbb{P}_{\theta,g}$, for any given g . Taylor's expansion theorem yields:

$$(B.2) \quad \forall k \in \mathbb{N} \quad \forall y \in \mathbb{R}_+ \quad \left| \underbrace{e^{-y} - \sum_{j=0}^{k-1} \frac{(-y)^j}{j!}}_{:=R_k(y)} \right| \leq \frac{|y|^k}{k!} \leq \frac{(e|y|)^k}{k^k}.$$

Thus, for all $z \in \mathcal{E}_R$, we have

$$\begin{aligned} \mathbb{P}_{\theta,g}(z) - \mathbb{P}_{\theta,\tilde{g}}(z) &= \pi^{-p} \int_0^1 e^{-\|z - \theta \bullet \varphi\|^2} \left[\frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi) \\ &= \pi^{-p} \sum_{j=0}^{k-1} \frac{(-1)^j}{j!} \int_0^1 \|z - \theta \bullet \varphi\|^{2j} \left[\frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi) \\ &\quad + \pi^{-p} \int_0^1 R_k(\|z - \theta \bullet \varphi\|^2) \left[\frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi). \end{aligned}$$

We now decompose $\theta = (\theta_{-\ell}, \dots, \theta_\ell)$ and $z = (z_{-\ell}, \dots, z_\ell)$ using polar coordinates: $\theta_m = \rho_m^{(1)} e^{i\alpha_m}$ and $z_m = \rho_m^{(2)} e^{i\beta_m}$ for $|m| \leq \ell$. This leads to

$$\|z - \theta \bullet \varphi\|^2 = \|z\|^2 + \|\theta\|^2 - 2 \sum_{m=-\ell}^{\ell} \rho_m^{(1)} \rho_m^{(2)} \cos(\beta_m - \alpha_m - m\varphi).$$

For any integer $j \leq k$, we deduce that

$$\|z - \theta \bullet \varphi\|^{2j} = C_j(z, \theta) + \sum_{r=1}^j \sum_{m=-\ell}^{\ell} a_{r,m}(z, \theta) [\cos(\beta_m - \alpha_m - m\varphi)]^r,$$

where $(a(r, m))_{r=1 \dots k, m=-\ell \dots \ell}$ is a complex matrix which only depends on z and θ . Using Euler's identity,

$$\|z - \theta \bullet \varphi\|^{2j} = C_j(z, \theta) + \sum_{r=-j\ell}^{j\ell} b_r(z, \theta) e^{ir\varphi},$$

where b stands for a complex vector obtained by the Binomial formula and coefficients $a_{r,m}(z, \theta)$. Consequently, for all $z \in \mathcal{E}_R$

$$\begin{aligned}
(\mathbb{P}_{\theta,g} - \mathbb{P}_{\theta,\tilde{g}})(z) &= \pi^{-p} \sum_{j=0}^{k-1} \frac{(-1)^j}{j!} \int_0^1 \left[C_j(z, \theta) \right. \\
&\quad \left. + \sum_{r=-j\ell}^{j\ell} b_r(z, \theta) e^{ir\varphi} \right] \left[\frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi) \\
&\quad + \pi^{-p} \int_0^1 R_k(\|z - \theta \bullet \varphi\|^2) \left[\frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi) \\
&= \pi^{-p} \sum_{j=0}^{k-1} \frac{(-1)^j}{j!} \left[C_j(z, \theta) c_0(g - \tilde{g}) + \sum_{r=-j\ell}^{j\ell} b_r(z, \theta) c_r(g - \tilde{g}) \right] \\
&\quad + \pi^{-p} \int_0^1 R_k(\|z - \theta \bullet \varphi\|^2) \left[\frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi).
\end{aligned}$$

Caratheodory's theorem shows that one can find \tilde{g} with a finite support of size $2(k-1)\ell + 1 \sim 2k\ell$ such that

$$\forall r \in [-(k-1)\ell, (k-1)\ell] \quad c_r(g) = c_r(\tilde{g}).$$

For such finite mixture law \tilde{g} , we obtain $\forall z \in \mathbb{C}^p$,

$$\mathbb{P}_{\theta,g}(z) - \mathbb{P}_{\theta,\tilde{g}}(z) = \pi^{-p} \int_0^1 R_k(\|z - \theta \bullet \varphi\|^2) \left[\frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi),$$

and of course

$$\begin{aligned}
(B) &\leq \pi^{-p} \int_{\mathcal{E}_R} \left| \int_0^1 R_k(\|z - \theta \bullet \varphi\|^2) \left[\frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi) \right| dz \\
&\leq 2\pi^{-p} \sup_{z \in \mathcal{E}_R, \varphi \in (0,1)} R_k(\|z - \theta \bullet \varphi\|^2) \text{Vol}(\mathcal{E}_R).
\end{aligned}$$

According to the choice $R = (1+a)\sqrt{(1+c)p}$ which implies that $\|z - \theta \bullet \varphi\| \leq (1+2a)\sqrt{(1+c)p}$, and using the volume of \mathcal{E}_R and Stirling's formula, we obtain

$$\begin{aligned}
(B) &\lesssim \pi^{-p} \frac{(e(1+2a)^2(1+c)p)^k}{k^k} \frac{\pi^p [(1+a)^2(1+c)p]^p}{p!} \\
&\lesssim C_1^p C_2^k e^{-k \log(k) + k \log(p)},
\end{aligned}$$

where we used in the last equation $p^p/p! \leq C^p$. If we define the threshold k in (B.2) such that $k \sim b\ell$ for a sufficiently large b , we then obtain for a universal C :

$$(B) = \int_{\mathcal{E}_R} |d\mathbb{P}_{\theta,g} - d\mathbb{P}_{\theta,\tilde{g}}|(z) \lesssim e^{\ell(C-b\log(b))}.$$

In order to bound (B) by $\epsilon/2$, we thus choose $k_\epsilon \sim b\ell_\epsilon$ for a sufficiently large absolute constant b . For such a choice, since $\log \frac{1}{\epsilon} \lesssim \ell_\epsilon$ we have found \tilde{g} with a discrete support of cardinal $s_\epsilon \sim 2b\ell_\epsilon^2$ points, with s_ϵ not depending on g , such that

$$d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{f,\tilde{g}}) \leq \epsilon/2.$$

Now, the first inequality in Proposition 3.3 comes from Proposition 3.2.

The second inequality in Proposition 3.3 is proved from the first one, using the relation $\|\theta\|_{\mathcal{H}_1} \leq \ell\|\theta\|$ valid for any $f \in \mathcal{H}_s^\ell$. \square

PROOF OF LEMMA 3.1. We follow a straightforward argument: $\mathbb{P}_{f,g}$ is a mixture model so

$$\mathbb{P}_{f,g} = \int_0^1 \mathbb{P}_{f,\delta_\alpha} dg(\alpha).$$

Thus

$$\begin{aligned} d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{\tilde{f},g}) &= \left\| \int_0^1 (\mathbb{P}_{f,\delta_\alpha} - \mathbb{P}_{\tilde{f},\delta_\alpha}) dg(\alpha) \right\|_{TV} \\ &\leq \int_0^1 \left\| \mathbb{P}_{f,\delta_\alpha} - \mathbb{P}_{\tilde{f},\delta_\alpha} \right\|_{TV} dg(\alpha) \\ &= \left\| \mathbb{P}_{f,\delta_0} - \mathbb{P}_{\tilde{f},\delta_0} \right\|_{TV} \leq d_H(\mathbb{P}_{f,\delta_0}, \mathbb{P}_{\tilde{f},\delta_0}). \end{aligned}$$

Assume now $Y \sim \mathbb{P}_{f,\delta_0}$, hence from (2.1) $dY = f(x)dx + dW$, with W is a complex standard Brownian motion. If we denote U a random variable $\mathcal{N}_{\mathbb{C}}(0,1)$, standard argument using Girsanov's formula yields

$$\begin{aligned} d_H^2(\mathbb{P}_{f,\delta_0}, \mathbb{P}_{\tilde{f},\delta_0}) &= 2 \left(1 - \mathbb{E}_{f,\delta_0} \sqrt{\frac{d\mathbb{P}_{\tilde{f},\delta_0}}{d\mathbb{P}_{f,\delta_0}}}(Y) \right) \\ &= 2 \left(1 - \mathbb{E}_{f,\delta_0} \sqrt{\exp(2\Re\langle \tilde{f} - f, dW \rangle - \|\tilde{f} - f\|^2)} \right) \\ &= 2 \left(1 - \exp\left(\frac{-\|\tilde{f} - f\|^2}{2}\right) \mathbb{E}_U \left[\exp(\|\tilde{f} - f\| \Re(U)) \right] \right) \\ &= 2 \left(1 - \exp\left(\frac{-\|\tilde{f} - f\|^2}{4}\right) \right) \leq \frac{\|\tilde{f} - f\|^2}{2}. \end{aligned}$$

□

B.2. Link between Kullback-Leibler and Hellinger neighbourhoods.

PROOF OF PROPOSITION 3.4. This proposition uses a corollary of Rice's formula (see [AW09] for various applications of such formula), stated in Lemma B.1 and postponed after this proof.

We begin with Girsanov's formula (3.1). Write now $Y = f^{0,-\tau} + W$ where W stands for a complex standard Brownian motion independent of the random shift τ (whose law is g^0). The L^2 norm is invariant with any shift thus

$$\begin{aligned} \frac{d\mathbb{P}_{f^0, g^0}}{d\mathbb{P}_{f, g}}(Y) &= \exp(\|f\|^2 - \|f^0\|^2) \frac{\int_0^1 e^{2\Re\langle f^{0,-\alpha_1}, f^{0,-\tau} + dW \rangle} dg^0(\alpha_1)}{\int_0^1 e^{2\Re\langle f^{-\alpha_2}, f^{0,-\tau} + dW \rangle} dg(\alpha_2)} \\ &\leq \exp(\|f\|^2 - \|f^0\|^2) \exp\left(2 \sup_{\alpha_1, \alpha_2} \Re\langle f^{0,-\alpha_1} - f^{-\alpha_2}, f^{0,-\tau} \rangle\right) \\ &\quad \exp\left(2 \sup_{\alpha_1, \alpha_2} \Re\langle f^{0,-\alpha_1} - f^{-\alpha_2}, dW \rangle\right) \\ &\leq e^{(\|f\| + \|f^0\|)^2} e^{Z_1 + Z_2}, \end{aligned}$$

where the last inequality is obtained using Cauchy-Schwarz's inequality and the notations

$$\begin{aligned} Z_1 &:= 2 \sup_{\alpha_1} \Re\langle f^{0,-\alpha_1}, dW \rangle = 2 \sup_{\alpha_1} \Re \int_0^1 \overline{f^0}(s - \alpha_1) dW_s, \\ Z_2 &:= 2 \sup_{\alpha_2} \Re\langle -f^{-\alpha_2}, dW \rangle = 2 \sup_{\alpha_2} \Re \int_0^1 -\overline{f}(s - \alpha_2) dW_s. \end{aligned}$$

We now set $\delta \in (0, 1]$ (it will be precisely fixed in the sequel) and we define the trajectories \mathcal{E}_δ as

$$\mathcal{E}_\delta := \left\{ Y = f^{0,-\tau} + W \quad \mid \quad \frac{d\mathbb{P}_{f^0, g^0}}{d\mathbb{P}_{f, g}}(Y) \geq e^{1/\delta} \right\}.$$

Hence, following the definition of M_δ^2 of (3.2), we have

$$M_\delta^2 = \mathbb{E}_{Y \sim \mathbb{P}_{f^0, g^0}} \left[\left(\frac{d\mathbb{P}_{f^0, g^0}}{d\mathbb{P}_{f, g}}(Y) \right)^\delta \mathbf{1}_{Y \in \mathcal{E}_\delta} \right].$$

For δ small enough, ($\delta \leq \frac{1}{2(\|f\| + \|f^0\|)^2}$):

$$\begin{aligned} M_\delta^2 &\leq e^{\delta(\|f\| + \|f^0\|)^2} \mathbb{E} e^{\delta(Z_1 + Z_2)} \mathbf{1}_{Z_1 + Z_2 \geq \frac{1}{\delta} - (\|f\| + \|f^0\|)^2} \\ &\leq e^{\delta(\|f\| + \|f^0\|)^2} \mathbb{E} e^{\delta(Z_1 + Z_2)} \mathbf{1}_{Z_1 + Z_2 \geq \frac{1}{2\delta}} \\ &\leq e^{\delta(\|f\| + \|f^0\|)^2} \mathbb{E} e^{\delta(Z_1 + Z_2)} \mathbf{1}_{e^{\delta(Z_1 + Z_2)} \geq \sqrt{e}}. \end{aligned}$$

Integrating by parts the last expectation, the use of Lemma B.1 yields

$$\begin{aligned} M_\delta^2 &\leq e^{\delta(\|f\| + \|f^0\|)^2} \int_{\sqrt{e}}^{+\infty} \mathbb{P} \left(e^{\delta(Z_1 + Z_2)} > u \right) du \\ &= e^{\delta(\|f\| + \|f^0\|)^2} \int_{\sqrt{e}}^{+\infty} \left[\mathbb{P} \left(\frac{Z_1}{2} \geq \frac{\log u}{4\delta} \right) + \mathbb{P} \left(\frac{Z_2}{2} \geq \frac{\log u}{4\delta} \right) \right] du \\ (B.3) \quad M_\delta^2 &\leq C(f^0, f) e^{\delta(\|f\| + \|f^0\|)^2} \int_{\sqrt{e}}^{+\infty} \left[e^{-\frac{\log^2(u)}{16\delta^2\|f^0\|^2}} + e^{-\frac{\log^2(u)}{16\delta^2\|f\|^2}} \right] du. \end{aligned}$$

Now, we can choose δ non negative and small enough such that $M_\delta^2 < \infty$ since for $u \geq \sqrt{e}$, we have

$$e^{-\frac{\log^2(u)}{16\delta^2\|f^0\|^2}} \leq e^{-\frac{\log(u)}{32\delta^2\|f^0\|^2}} = u^{-1/32\delta^2\|f^0\|^2},$$

which is an integrable function as soon as $\delta^2 < \frac{1}{32\|f^0\|^2}$, and the same holds with f instead of f_0 . Note that M_δ^2 is uniformly bounded if f is picked into a ball centered at 0 with radius $2\|f^0\|$. \square

We now show that the technical inequality used in (B.3) is satisfied.

LEMMA B.1. *Let W a complex standard Brownian motion and u a complex 1-periodic map of \mathcal{H}_s . We assume that u is of class \mathcal{C}^2 . Then when $t/\|u\| \rightarrow +\infty$, we have*

$$\mathbb{P} \left(\sup_{\alpha} \Re \langle u^{-\alpha}, dW \rangle > t \right) \lesssim \frac{\|u'\|}{2\pi\|u\|} \exp \left(\frac{-t^2}{\|u\|^2} \right).$$

In particular, if $u \in \mathcal{H}_s^\ell$, we have

$$\mathbb{P} \left(\sup_{\alpha} \Re \langle u^{-\alpha}, dW \rangle > t \right) \lesssim \frac{\ell}{2\pi} \exp \left(\frac{-t^2}{\|u\|^2} \right).$$

PROOF. We define the following process

$$\forall \alpha \in [0, 1] \quad X(\alpha) := \frac{\sqrt{2} \Re \left(\int_0^1 \bar{u}(s - \alpha) dW_s \right)}{\|u\|}.$$

X is a Gaussian centered process. Its covariance function is given by

$$\Gamma(t) = \mathbb{E}[X(0)X(t)].$$

Obviously, one has $\Gamma(0) = 1$ and Cauchy-Schwarz's inequality implies that $\Gamma(s) \leq \Gamma(0)$. Moreover, since Γ is $\mathcal{C}^1([0, 1])$, we deduce that $\Gamma'(0) = 0$ and simple computation yields

$$\Gamma''(0) = \frac{\Re\left(\int_0^1 u'(s)u''(s)ds\right)}{\|u\|^2} = -\frac{\|u'\|^2}{\|u\|^2}.$$

Rice's formula (see for instance exercise 4.2, chapter 4 of [AW09]) then yields that when $t \rightarrow +\infty$, we have

$$\mathbb{P}\left(\sup_{\alpha} X(\alpha) > t\right) \sim \frac{\|u'\|}{2\pi\|u\|} e^{-t^2/2}.$$

This ends the proof of the first inequality. Assume furthermore that $u \in \mathcal{H}_s^\ell$, Parseval's equality implies that $\|u'\| \leq \ell\|u\|$ and we obtain the second inequality. \square

B.3. Hellinger neighbourhoods.

PROOF OF PROPOSITION 3.5. Recall first that if Y follows \mathbb{P}_{f^0, g^0} , one shift β is randomly sampled according to g^0 . Conditionally to this shift β , Y is described through a white noise model $dY(x) = f^0(x - \beta)dx + dW(x)$. For any function F of the trajectory Y , we will denote $\mathbb{E}_\beta F(Y)$ the expectation of $F(Y)$ up to the condition that the shift is equal to β , and of course one has

$$\mathbb{E}_0[F(Y)] = \int_0^1 \mathbb{E}_\beta[F(Y)]dg^0(\beta).$$

For each possible value of $\beta \in [0, 1]$, we define

$$\begin{aligned} D_\beta(\alpha) &:= \exp\left(2\Re\langle f_{\ell_n}^{0, -\alpha}, f^{0, -\beta} \rangle + 2\Re\langle f_{\ell_n}^{0, -\alpha}, dW \rangle - \|f_{\ell_n}^0\|^2\right), \\ X_\beta(\alpha) &:= \exp\left(2\Re\langle (f^0 - f_{\ell_n}^0)^{-\alpha}, f^{0, -\beta} \rangle \right. \\ &\quad \left. + 2\Re\langle (f^0 - f_{\ell_n}^0)^{-\alpha}, dW \rangle - \|f^0 - f_{\ell_n}^0\|^2\right). \end{aligned}$$

We can now split the randomness of the Brownian motion into two parts: the first one is spanned by the Fourier frequencies from $-\ell_n$ to ℓ_n and the second part is its orthogonal (in L^2): $W = W_1 + W_2$. Of course, W_1 and W_2 are independent.

Moreover, $\langle f_{\ell_n}^{0,-\alpha}, dW \rangle = \langle f_{\ell_n}^{0,-\alpha}, dW_1 \rangle$ and $\langle (f^0 - f_{\ell_n}^0)^{-\alpha}, dW \rangle = \langle (f^0 - f_{\ell_n}^0)^{-\alpha}, dW_2 \rangle$. For any fixed β , $D_\beta(\alpha)$ is measurable with respect to the filtration associated to W_1 , and $X_\beta(\alpha)$ is independent of W_1 . We thus obtain using Jensen's inequality and this filtration property that

$$\begin{aligned} (\tilde{E}_1)^2 &= \mathbb{E} \left[\log \frac{\int_0^1 D_\beta(\alpha) X_\beta(\alpha) dg^0(\alpha)}{\int_0^1 D_\beta(\alpha) dg^0(\alpha)} \right] \\ &\leq \log \int_0^1 \mathbb{E}_\beta^{W_2} \left[\mathbb{E}_\beta^{W_1} \left[\frac{\int_0^1 D_\beta(\alpha) X_\beta(\alpha) dg^0(\alpha)}{\int_0^1 D_\beta(\alpha) dg^0(\alpha)} \middle| W_2 \right] \right] dg^0(\beta) \\ &\leq \log \int_0^1 \mathbb{E}_\beta^{W_2} \left[X_\beta(\alpha) \mathbb{E}_\beta^{W_1} \left[\frac{\int_0^1 D_\beta(\alpha) dg^0(\alpha)}{\int_0^1 D_\beta(\alpha) dg^0(\alpha)} \middle| W_2 \right] \right] dg^0(\beta) \\ &\leq \log \int_0^1 \left(\sup_\alpha \mathbb{E}_\beta^{W_2} [X_\beta(\alpha)] \right) dg^0(\beta). \end{aligned}$$

The notation $\mathbb{E}_\beta^{W_1} F(Y)$ (resp. $\mathbb{E}_\beta^{W_2} F(Y)$) used above refers to the expectation of $F(Y)$ with respect to W_1 (resp. with respect to W_2) with a fixed β .

Now, one should remark that $X_\beta(\alpha)$ has the same law as

$$\exp \left(2\Re \langle (f^0 - f_{\ell_n}^0)^{-\alpha}, f^{0,-\beta} \rangle + U \right),$$

where $U \sim \mathcal{N}_{\mathbb{R}}(-\|f^0 - f_{\ell_n}^0\|^2, 2\|f^0 - f_{\ell_n}^0\|^2)$, and $\mathbb{E}[e^U] = 1$. Hence

$$\begin{aligned} (\tilde{E}_1)^2 &\leq \log \int_0^1 \sup_\alpha \exp \left(2\Re \langle (f^0 - f_{\ell_n}^0)^{-\alpha}, f^{0,-\beta} \rangle \right) dg^0(\beta) \\ &\leq \log \sup_{\alpha, \beta} \exp \left(2\Re \langle (f^0 - f_{\ell_n}^0)^{-\alpha}, f^{0,-\beta} \rangle \right) \end{aligned}$$

We can now switch log and sup since log is increasing, and we obtain

$$(\tilde{E}_1) \leq \sqrt{2 \sup_{\alpha, \beta} \Re \langle (f^0 - f_{\ell_n}^0)^{-\alpha}, f^{0,-\beta} \rangle}.$$

Again, we can use the orthogonal decomposition $f^{0,-\beta} = f_{\ell_n}^{0,-\beta} + f^{0,-\beta} - f_{\ell_n}^{0,-\beta}$ and Cauchy-Schwarz's inequality yields $(\tilde{E}_1) \leq \sqrt{2} \|f^0 - f_{\ell_n}^0\|$.

Note that untill now we did not use the hypothesis $f^0 \in \mathcal{H}_s$. It is only needed to get the last inequality in Proposition 3.5. \square

To establish Lemma 3.2, we first remind the following useful result.

LEMMA B.2. *For any any dimension p and any couple of points $(z_1, z_2) \in \mathbb{C}^p$, if $\|z_1 - z_2\|$ is the Euclidean distance in \mathbb{C}^p , then one has*

$$d_{TV}(\gamma_{z_1}, \gamma_{z_2}) = \frac{1}{2} \|\gamma_{z_1} - \gamma_{z_2}\|_{L_1} = \left[2\Phi\left(\frac{\|z_1 - z_2\|}{2}\right) - 1 \right] \leq \frac{\|z_1 - z_2\|}{\sqrt{2\pi}},$$

where Φ stands for the cumulative distribution function of a real standard Gaussian variable.

PROOF OF LEMMA 3.2. Adapting the proof of Lemma 5.1 of [GvdV01],

$$\begin{aligned} \|\mathbb{P}_{f_{\ell_n}^0, \tilde{g}} - \mathbb{P}_{f_{\ell_n}^0, \tilde{g}}\|_{L_1} &\leq \sum_{j=1}^J \int_{\varphi_j - \eta/2}^{\varphi_j + \eta/2} \|\gamma_0(\cdot - \theta \bullet \varphi) - \gamma_0(\cdot - \theta \bullet \varphi_j)\|_{L_1} d\tilde{g}(\varphi) \\ &\quad + 2 \sum_{j=1}^J |\tilde{g}([\varphi_j - \eta/2, \varphi_j + \eta/2]) - p_j|. \end{aligned}$$

Using Lemma B.2 ends the proof. \square

PROOF OF PROPOSITION 3.7. The construction used in the proof of Proposition 3.3 provide a mixture \tilde{g} such that \tilde{g} is supported by $\tilde{J}_n := C\ell_n^2$ points (denoted $(\varphi_j)_{j=1 \dots \tilde{J}_n}$) so that $d_H(\mathbb{P}_{f_{\ell_n}^0, g^0}, \mathbb{P}_{f_{\ell_n}^0, \tilde{g}}) \leq \epsilon_n$. Therefore $\tilde{g} = \sum_{j=1}^{\tilde{J}_n} w_j \delta_{\varphi_j}$. As pointed by [GvdV01], one can slightly modify \tilde{g} so that the support points are separated enough as follows. First, denote $(\psi_j)_{j=1 \dots J_n}$ the subset of $(\varphi_j)_{j=1 \dots \tilde{J}_n}$ which is η_n -separated with a maximal number of elements. Hence, $J_n \leq \tilde{J}_n$ and up to a permutation, one can divide $(\varphi_j)_{j=1 \dots \tilde{J}_n}$ in two parts: $(\varphi_j)_{j=1 \dots \tilde{J}_n} = (\psi_j)_{j=1 \dots J_n} \cup (\varphi_j)_{j=J_n+1 \dots \tilde{J}_n}$. For any $i \in \{J_n + 1, \dots, \tilde{J}_n\}$, we define $\psi_{j(i)}$ as the closest point of $(\psi_j)_{j=1 \dots J_n}$, the new discrete mixture law is then given by

$$\tilde{g} = \sum_{j=1}^{J_n} \underbrace{\left(w_j + \sum_{i > J_n | j(i)=j} w_i \right)}_{:= \tilde{w}_j} \delta_{\psi_j}.$$

Of course, \tilde{g} as a support which is η_n -separated. Moreover, we have

$$\begin{aligned}
2d_{TV}(\mathbb{P}_{f_{\ell_n}^0, \tilde{g}}, \mathbb{P}_{f_{\ell_n}^0, \tilde{\tilde{g}}}) &= \int_{\mathbb{C}^{\ell_n}} \left| \sum_{i=1}^{J_n} \tilde{w}_i \gamma(z - \theta \bullet \psi_i) - \sum_{i=1}^{\tilde{J}_n} w_i \gamma(z - \theta \bullet \varphi_i) \right| dz \\
&= \int_{\mathbb{C}^{\ell_n}} \left| \sum_{j=1}^{J_n} (\tilde{w}_j - w_j) \gamma(z - \theta \bullet \psi_j) - \sum_{i>J_n} w_i \gamma(z - \theta \bullet \varphi_i) \right| dz \\
&= \int_{\mathbb{C}^{\ell_n}} \left| \sum_{j=1}^{J_n} \sum_{i>J_n | j(i)=j} w_i [\gamma(z - \theta \bullet \psi_j) - \gamma(z - \theta \bullet \varphi_i)] \right| dz.
\end{aligned}$$

Then, Fubini's theorem yields

$$d_{TV}(\mathbb{P}_{f_{\ell_n}^0, \tilde{g}}, \mathbb{P}_{f_{\ell_n}^0, \tilde{\tilde{g}}}) \leq \sum_{j=1}^{J_n} \sum_{i>J_n | j(i)=j} w_i d_{TV}(\gamma_{\theta \bullet \varphi_i}, \gamma_{\theta \bullet \psi_j}),$$

and we deduce from Lemma B.2 that

$$d_{TV}(\mathbb{P}_{f_{\ell_n}^0, \tilde{g}}, \mathbb{P}_{f_{\ell_n}^0, \tilde{\tilde{g}}}) \leq \sqrt{2\pi} \sum_{j=1}^{J_n} \sum_{i>J_n | j(i)=j} w_i \|\theta\|_{\mathcal{H}_1} \eta_n \leq \sqrt{2\pi} \|\theta\|_{\mathcal{H}_1} \eta_n.$$

Now the relations between Hellinger and Total Variation distances (A.2) yield

$$d_H(\mathbb{P}_{f_{\ell_n}^0, g^0}, \mathbb{P}_{f_{\ell_n}^0, \tilde{g}}) \leq \epsilon_n + d_H(\mathbb{P}_{f_{\ell_n}^0, \tilde{g}}, \mathbb{P}_{f_{\ell_n}^0, \tilde{\tilde{g}}}) \leq \left(1 + (8\pi)^{1/4} \|\theta\|_{\mathcal{H}_1}^{1/2}\right) \epsilon_n.$$

Lemma 3.2 permits to conclude. \square

B.4. Checking the conditions of Theorem 2.1.

PROOF OF PROPOSITION 4.1. We have seen in the proof of Proposition 3.4 that M_δ^2 is uniformly bounded with respect to $\|f\|$ and $\|f^0\|$ for a suitable choice of δ . We restrict our study to the elements f such that $\|f\| \leq 2\|f^0\|$.

We know from Proposition 3.4 and Theorem 3.2 that as soon as $\tilde{\epsilon}_n \log \frac{1}{\tilde{\epsilon}_n} \leq c\epsilon_n$ with c small enough:

$$\begin{aligned}
\mathcal{V}_{\tilde{\epsilon}_n}(\mathbb{P}_{f^0, g^0}, d_H) &:= \{\mathbb{P}_{f, g} \in \mathcal{P} | d_H(\mathbb{P}_{f^0, g^0}, \mathbb{P}_{f, g}) \leq \tilde{\epsilon}_n \text{ and } \|f\| \leq 2\|f^0\|\} \\
&\subset \mathcal{V}_{\epsilon_n}(\mathbb{P}_{f^0, g^0}, d_{KL}).
\end{aligned}$$

This last condition on $\tilde{\epsilon}_n$ is true as soon as

$$(B.4) \quad \tilde{\epsilon}_n := \tilde{c}\epsilon_n \left(\log \frac{1}{\epsilon_n} \right)^{-1}$$

with \tilde{c} small enough. Now, Proposition 3.8 permits to describe a subset of $\mathcal{V}_{\tilde{\epsilon}_n}(\mathbb{P}_{f^0, g^0}, d_H)$, by the definition of subsets $\mathcal{F}_{\tilde{\epsilon}_n}$ and $\mathcal{G}_{\tilde{\epsilon}_n}$ for f and g . Choose $\ell_n := \tilde{\epsilon}_n^{-1/s}$.

We first bound the prior mass on $\mathcal{G}_{\tilde{\epsilon}_n}$. This follows from the lower bound given by Lemma 4.1. The prior for g is a Dirichlet process with a finite base measure α admitting a continuous positive density on $[0, 1]$. Since η_n goes to zero, for n large enough $\alpha(\psi_j - \eta_n/2, \psi_j + \eta_n/2)$ for any $j = 1 \dots J_n$. Note that $J_n \lesssim \ell_n^2 = \tilde{\epsilon}_n^{-2/s} \leq \tilde{\epsilon}_n^{-2}$. Thus, there exists an absolute constant $a \in (0, 1]$ such that the condition $J_n \leq 2(a\tilde{\epsilon}_n)^{-2}$ is fulfilled, and one can find universal constants C and c such that for n large enough

$$(B.5) \quad \Pi_n(\mathcal{G}_{\tilde{\epsilon}_n}) \geq \Pi_n(\mathcal{G}_{a\tilde{\epsilon}_n}) \geq Ce^{-cJ_n \log \frac{1}{\tilde{\epsilon}_n}} \geq Ce^{-c\ell_n^2 \log \frac{1}{\tilde{\epsilon}_n}}.$$

We next consider the prior mass on $\mathcal{F}_{\tilde{\epsilon}_n}$. Remark that when n is large enough, any element of $\mathcal{F}_{\tilde{\epsilon}_n}$ satisfies $\|f\| \leq 2\|f^0\|$ and the additional condition on $\|f\|$ in the definition of $\mathcal{V}_{\tilde{\epsilon}_n}(\mathbb{P}_{f^0, g^0}, d_H)$ is instantaneously fulfilled. Remark that from the construction of our prior on f , one has

$$\Pi_n(\mathcal{F}_{\tilde{\epsilon}_n}) \geq \lambda(\ell_n) \times \pi_{\ell_n}(B(\theta_{\ell_n}^0, \tilde{\epsilon}_n^2)).$$

From our assumption on the prior λ , we have $\lambda(\ell_n) \geq e^{-c\ell_n^2 \log^\rho \ell_n}$, and the value of the volume of the $(4\ell_n + 2)$ -dimensional Euclidean ball of radius $\tilde{\epsilon}_n^2$ implies

$$\Pi_n(\mathcal{F}_{\tilde{\epsilon}_n}) \geq e^{-c\ell_n^2 \log^\rho \ell_n} \inf_{u \in B(0, \tilde{\epsilon}_n^2)} \left(\frac{e^{-\|\theta^0 + u\|^2 / \xi_n^2}}{\pi^{2\ell_n + 1} \xi_n^{2(2\ell_n + 1)}} \right) (\tilde{\epsilon}_n^2)^{4\ell_n + 2} \frac{\pi^{2\ell_n + 1}}{\Gamma(2\ell_n + 2)}.$$

For n large enough we get

$$(B.6) \quad \begin{aligned} \Pi_n(\mathcal{F}_{\tilde{\epsilon}_n}) &\geq \exp - [c\ell_n^2 \log^\rho \ell_n + \xi_n^{-2} \\ &\quad + (2\ell_n + 1)(\log \ell_n + 4 \log(1/\tilde{\epsilon}_n) - \log \xi_n^{-2} + \mathcal{O}(1))] \\ &\geq \exp [-(c + o(1)) [\ell_n^2 \log^\rho \ell_n \vee \xi_n^{-2}]] \end{aligned}$$

Gathering (B.5) and (B.6), the relations $\ell_n = \tilde{\epsilon}_n^{-1/s}$ and (B.4) lead to

$$\begin{aligned} \Pi_n(\mathcal{V}_{\epsilon_n}(\mathbb{P}_{f^0, g^0}, d_{KL})) &\geq \Pi_n(\mathcal{F}_{\tilde{\epsilon}_n}) \Pi_n(\mathcal{G}_{\tilde{\epsilon}_n}) \\ &\geq \exp\left[-(c + o(1)) \left[\ell_n^2 \log^\rho \ell_n \vee \xi_n^{-2}\right]\right] \\ &\geq \exp\left[-(c + o(1)) \left[\tilde{\epsilon}_n^{-2/s} \log^\rho(1/\tilde{\epsilon}_n) \vee \xi_n^{-2}\right]\right] \\ &\geq \exp\left[-(c + o(1)) \left[\epsilon_n^{-2/s} (\log(1/\epsilon_n))^{\rho+2/s} \vee \xi_n^{-2}\right]\right] \end{aligned}$$

for constants $c > 0$. \square

PROOF OF PROPOSITION 4.2. The upper bound on the packing number comes directly from Theorem 3.1 since we set $w_n = \sqrt{2k_n + 1}$.

Now, to control the prior mass outside the sieve, remark first that owing to the construction of our prior, we have

$$(B.7) \quad \Pi_n(\mathcal{P} \setminus \mathcal{P}_{k_n, w_n}) \leq \sum_{|k| \geq k_n} \lambda(k) + \Pr\left(\sum_{|k| \leq k_n} |\theta_k|^2 \geq w_n^2\right),$$

where each θ_k for $-k_n \leq k \leq k_n$ follows a centered Gaussian law of variance ξ_n^2 . Now, there exists some constants c and C such that for sufficiently large n :

$$\sum_{|k| \geq k_n} \lambda(k) \leq C\lambda(k_n) \leq e^{-ck_n^2 \log^\rho(k_n)}.$$

Regarding now the second term of the upper bound in (B.7), we use (B.1) to get

$$\begin{aligned} \Pr\left(\sum_{|k| \leq k_n} |\theta_k|^2 \geq w_n^2\right) &= \Pr\left(\sum_{|k| \leq k_n} \left|\frac{\theta_k}{\xi_n}\right|^2 \xi_n^2 \geq w_n^2\right) \\ &\leq \mathbb{P}\left(\chi_{2k_n+1}^2 \geq 2(2k_n+1)\xi_n^{-2}\right) \\ &\leq \frac{1}{(\xi_n^{-2} - 1)\sqrt{\pi}} e^{-(2k_n+1)[\xi_n^{-2} - 1 - \log \xi_n^{-2}] - \log(2k_n+1)/2}. \end{aligned}$$

Now, using the value of ξ_n , we obtain

$$\Pi_n(\mathcal{P} \setminus \mathcal{P}_{k_n, w_n}) \leq e^{-c[k_n^2 \log^\rho(k_n) \wedge k_n \xi_n^{-2}]}.$$

This concludes the proof of the Proposition. \square

ACKNOWLEDGEMENTS

S. G. is indebted to Jean-Marc Azaïs for stimulating discussions related to some technical parts of this work. Authors also thank Jérémie Bigot, Ismaël Castillo, Xavier Gendre, Judith Rousseau and Alain Trouvé for enlightening exchanges.

REFERENCES

- [AAT07] S. Allasonnière, Y. Amit, and A. Trouvé. Toward a coherent statistical framework for dense deformable template estimation. *Journal of the Royal Statistical Society (B)*, 69:3–29, 2007.
- [AGP91] Y. Amit, U. Grenander, and M. Piccioni. Structural Image Restoration through Deformable Templates. *JASA*, 86(414):376–387, 1991.
- [AKT10] Stéphanie Allasonnière, Estelle Kuhn, and Alain Trouvé. Construction of Bayesian deformable models via a stochastic approximation algorithm: a convergence study. *Bernoulli*, 16(3):641–678, 2010.
- [AW09] Jean-Marc Azaïs and Mario Wschebor. *Level sets and extrema of random processes and fields*. John Wiley & Sons Inc., Hoboken, NJ, 2009.
- [BCG12] J. Bigot, C. Christophe, and S. Gadat. Random action of compact Lie groups and minimax estimation of a mean pattern. *IEEE, Transactions on Information Theory*, 58(6):3509–3520, 2012.
- [BCK07] R. Blundell, X. Chen, and D. Kristense. Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves. *Econometrica*, 75(6):1613–1669, 2007.
- [BG10] Jérémie Bigot and Sébastien Gadat. A deconvolution approach to estimation of a common shape in a shifted curves model. *Ann. Statist.*, 38(4):2422–2464, 2010.
- [BG12] Jérémie Bigot and Xavier Gendre. Minimax properties of Fréchet means of discretely sampled curves. *Preprint*, 2012.
- [BG13] D. Bontemps and S. Gadat. Bayesian methods in the Shape Invariant Model (II): Identifiability and posterior consistency on functional spaces. *Preprint*, pages x–x+y, 2013.
- [BGKM12] Jérémie Bigot, Sébastien Gadat, Thierry Klein, and Clément Marteau. Intensity estimation of non-homogeneous Poisson processes from shifted trajectories. *Preprint*, 2012.
- [BGL09] J. Bigot, S. Gadat, and J.M. Loubes. Statistical M-Estimation and Consistency in large deformable models for Image Warping. *Journal of Mathematical Imaging and Vision*, 34(3):270–290, 2009.
- [BGV09] J. Bigot, F. Gamboa, and M. Vimond. Estimation of translation, rotation and scaling between noisy images using the Fourier Mellin transform. *SIAM Journal on Imaging Sciences*, 2(2):614–645, 2009.
- [Big06] J. Bigot. Landmark-based registration of curves via the continuous wavelet transform. *Journal of Computational and Graphical Statistics*, 15(3):542–564, 2006.
- [Big11] Jérémie Bigot. Fréchet means of curves for signal averaging and application to ECG data analysis. *preprint*, 2011.
- [BSW99] A. Barron, M. J. Schervish, and L. Wasserman. The consistency of posterior distributions in nonparametric problems. *Annals of Statistics*, 27:536–561, 1999.

- [Cas12] Ismaël Castillo. A semi-parametric Bernstein-von Mises theorem for Gaussian process priors. *Probability Theory and Related Fields*, 1:53–99, 2012.
- [CD11] O. Collier and A Dalalyan. Curve registration by nonparametric goodness-of-fit testing. *Preprint*, 2011.
- [Col12] Olivier Collier. Minimax hypothesis testing for curve registration. *Electronic Journal of Statistics*, 6:1129–1154, 2012.
- [Fer73] Thomas S. Ferguson. A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, 1:209–230, 1973.
- [GGvdV00] Subhashis Ghosal, Jayanta K. Ghosh, and Aad W. van der Vaart. Convergence rates of posterior distributions. *Ann. Statist.*, 28(2):500–531, 2000.
- [GM01] C. A. Glasbey and K. V Mardia. A penalized likelihood approach to image warping. *Journal of the Royal Statistical Society (B)*, 63(3):465–514, 2001.
- [GM07] U. Grenander and M. Miller. *Pattern Theory: From Representation to Inference*. Oxford Univ. Press, Oxford, 2007.
- [GvdV01] Subhashis Ghosal and Aad W. van der Vaart. Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.*, 29(5):1233–1263, 2001.
- [GW00] Christopher R. Genovese and Larry Wasserman. Rates of convergence for the Gaussian mixture sieve. *Ann. Statist.*, 28(4):1105–1127, 2000.
- [IH81] I. A. Ibragimov and R. Z. Has’minskii. *Statistical estimation*, volume 16 of *Applications of Mathematics*. Springer-Verlag, New York, 1981. Asymptotic theory.
- [IL06] Tadeusz Inglot and Teresa Ledwina. Asymptotic optimality of new adaptive test in regression model. *Ann. Inst. H. Poincaré Probab. Statist.*, 42(5):579–590, 2006.
- [KG88] A. Kneip and T. Gasser. Convergence and consistency results for self-modelling regression. *Ann. Statist.*, 16:82–112, 1988.
- [LCY00] Lucien Le Cam and Grace Yang. *Asymptotics in Statistics: Some Basic Concepts*. Springer series in statistics. Springer Verlag, New-York, 2000.
- [MM11] Cathy Maugis and Bertrand Michel. A non asymptotic penalized criterion for Gaussian mixture model selection. *ESAIM Probab. Statist.*, 15:41–68 (electronic), 2011.
- [MMW07] J. Mortazavi, R. Myers, and B. Wold. Genome-wide mapping of in vivo protein- dna interactions. *Science*, 316:1497–1502, 2007.
- [PMRC10] Wooram Park, Dean R. Madden, Daniel N. Rockmore, and Gregory S. Chirikjian. Deblurring of class-averaged images in single-particle electron microscopy. *Inverse Problems*, 26(3):035002, 29, 2010.
- [RR12] Vincent Rivoirard and Judith Rousseau. Posterior concentration rates for infinite dimensional exponential families. *Bayesian Analysis*, 7(2):311–334, 2012.
- [WS95] Wing Hung Wong and Xiaotong Shen. Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.*, 23(2):339–362, 1995.

INSTITUT MATHÉMATIQUES DE TOULOUSE, UNIVERSITÉ PAUL SABATIER
 118 ROUTE DE NARBONNE F-31062 TOULOUSE CEDEX 9 FRANCE
 E-MAIL: dominique.bontemps@math.univ-toulouse.fr
 E-MAIL: sebastien.gadat@math.univ-toulouse.fr